

# Arms talks enter muddy waters

This week's negotiations on conventional forces in Europe will be more complicated than a few years ago, paradoxically because Mr Gorbachev has changed the agenda.

THE principle that any negotiations on arms control are better than none may be falsified by the talks begun this week in Vienna on the reduction of conventional forces in Europe. Ordinarily it would be a hopeful sign that the Warsaw Pact and the North Atlantic Treaty Organisation (NATO) are tackling directly the most serious threat to their mutual security — the risk of a European conflict — but that is only half the story. It is more important that neither side has a clear vision of its objectives, while each has taken an opening position which is certain not to be accepted by the other. So there are two dangers. One, the lesser, is that the negotiations will lapse into the ill-tempered stalemate of the talks on Mutually Balanced Force Reductions (MBFR) which dragged on at Vienna for nearly a decade until they were abandoned last year. The second is that they will blow up in the participants' faces, undermining the amity of the past two years.

The Soviet position, dramatized by Mr Mikhail Gorbachev's declaration at the United Nations of a unilateral reduction of conventional forces, is that there should also be a demilitarized zone in Central Europe, on either side of the mutual German frontier. The obvious objection is that such an agreement would put NATO at a serious disadvantage. The distance between the German frontier and the Rhine is already so small, especially in NATO's northern sector, that the effective defence of northern West Germany is an insomniac's conundrum. In any case, NATO says, the volunteered reductions of battle tanks will still leave the Warsaw Pact with more of them than the limit of 20,000 which is the centrepiece of its own opening bid. There is plenty of room for argument, but no obvious compromise, in all that.

NATO's starting position, which would exclude both battlefield nuclear weapons and aircraft from the negotiations, is similarly inconsistent. But nuclear weapons, for example, are said to be a necessary compensation for the Warsaw Pact's superiority in battle tanks. If there were an incursion from the East, NATO would attempt to halt it by using tactical nuclear weapons. So is it not fair that the Warsaw Pact should ask that NATO should trade its dependence on nuclear interdiction for an agreement that there should be parity on battle tanks? The same applies to aircraft, which can be used in interdiction as well as in other roles.

There are more serious difficulties with which neither of the participants has come to grips, but which cannot be

settled at Vienna. Part of NATO's opening gambit is that no more than 30 per cent of either side's quota of forces should be provided by a single nation. That is a way of teasing the Soviet Union about the military independence of its allies in Eastern Europe, but it also begs serious questions that the Soviet Union cannot answer now, so long as its allies have not had time to figure out what version of *perestroika* they will follow. The same is true of NATO's own little local difficulty, the West German government's reluctance to agree to the deployment of an improved version of the Lance short-range nuclear missile.

The big boys, the Soviet Union and the United States, should recognize that they cannot know what might best emerge from their negotiations until the political future of Central Europe has been clarified. Meanwhile, if ambition does not get the better of good sense, there is much that can be done. The negotiators could usefully begin where they would have to finish, by deciding how to verify agreed quotas of tanks and other battlefield machines. The framework of an agreement to pull back battlefield nuclear weapons from the interface would be useful, if not for now, then for the future. Trying to distinguish between offensive and defensive arms would be beneficial. But a lasting agreement at Vienna will require a political understanding that cannot be quickly reached. Better to recognize that at the outset than to spoil the future by hurrying. □

## Tokyo's brave reform

The University of Tokyo is embarking on a daring reform — not before time.

THE University of Tokyo may still be just a little ahead of the University of Kyoto in the affections of young Japanese seeking a university education, but, sadly, neither ranks with the great universities of the West as an international centre of scholarship. There are scholars with international reputations at both places, of course, but that is not the same thing. Even the best of Japan's national universities are parochial and static institutions compared with the cosmopolitan intellectual transit-camps which are not merely the custodians but also the sources of scholarship elsewhere.

Now, the University of Tokyo, or at least its science faculty, plans to change. There has been talk of reform for



several years, much of it prompted by Professor Akito Arima, the newly elected president (see *Nature* 338, 8; 1989). Now, a committee under Professor Akiyoshi Wada, the physicist turned molecular biologist, has put forward a plan to reorganize the graduate school of the science faculty. The plan would transform the education of scientists at Tokyo, but that is only half of it. The true objective is to create a science faculty which is also an internationally connected research community. Wada, who becomes the dean of the graduate school on 1 April, will be well-placed, but success will hang on whether the Japanese government will endow the plan with resources — and turn a deaf ear to the grumbles there will be that Tokyo is seeking merely to perpetuate its present exalted status and to fight off moves by the national laboratories to run their own graduate programmes.

The oddest attribute of modern Japan is the ostentatiously un-modern condition of its universities. The curriculum is rigid; two years of general education, then two of specialized education (including some experience of working in a research laboratory), for every undergraduate. The structure of the faculty is similarly cast in concrete; the *koza* system attaches to each full professor (*kyoju*), who is in himself an autonomous academic unit, a couple of more junior academics and a research assistant who may become, if that is the great man's inclination (there are few women professors), his intellectual skivvies. (However menial their function and however meagre their salaries, all these people have tenure.) There is also a portion of the university's disposable income to spend on research, usually so small that it would seem derisory even to Britain's pauperized academics. Even at the University of Tokyo, there are buildings so shabby, ill-equipped and dust-ridden that they seem hardly to have been touched since they were built, mostly in the rebuilding programme following the great earthquake and fire of 1923.

How, in these circumstances, can research flourish? That many Japanese academics have won glowing international reputations is a testament to the durability and ingenuity of the human spirit. In the Japanese system as it is, the pursuit of understanding is but foolish self-indulgence. Doing nothing is much easier, but the simplest way to a seemingly academic life is for a *kyoju* to strike a few bargains with industrial companies, which will bring decent equipment, and to train graduate students on related applied research (ensuring that they also have comfortable jobs). Exceptional circumstances do arise, as when a bunch of physicists win approval (and funds for) a space programme, or when a government laboratory is sited on a university campus, but these merely show how moribund and inward-looking is the system as now constructed.

The new recipe for Tokyo's science faculty would change this from the bottom up. First, most undergraduates after the two years of general education would set out on a four- or five-year course leading to a research degree. A crucial element of the plan is that those concerned

should work in basic research — which is where the rub lies. For the plan will succeed only if there are enough more senior people practising interesting research with whom students can rub shoulders. That argues for what has long been Japan's most urgent academic need, more positions at the postdoctoral level than the few hundred fellowships now administered by the Japan Society for the Promotion of Science. There will also need to be research funds.

In the long run, only the Ministry of Education can ensure that Tokyo's plan succeeds. Does it fully appreciate the importance of the venture? The declared objective is to make scholarship seem an honourable pursuit and to create an intellectual climate that will be attractive to people from overseas. (While all academic posts at the national universities are in principle open to foreigners, only a tiny proportion are thus occupied.) Japan does not have to swallow the complaint by US diplomats that the monoglot character of the national universities is proof of yet another restraint of trade to know that the richest country in the world should hang its head when even its grandest universities are such poor magnets for scholars from elsewhere. □

## Convictions on ozone

Mrs Margaret Thatcher's crusade in favour of the ozone layer should not blind her to greenhouse gases.

Now that the British Prime Minister's zeal is harnessed to the preservation of the ozone layer, the rest of us will be able to sleep more soundly when we sunbathe. That is one of the lessons to be drawn from the conference in London at the weekend (see page 101), when Mrs Thatcher recruited a further score of signatures for the Montreal Protocol which commits its signatories to a 50 per cent reduction of eight chemicals believed to affect the ozone layer. That is a considerable achievement for the persuasiveness of conviction, but there is a long way to go.

This journal's conviction is that the Montreal Protocol is chiefly to be valued because it is a precedent for the much more constraining convention there will have to be if ever it becomes necessary to restrict by international agreement the emission of carbon dioxide so as to avoid the now-familiar greenhouse effect. It is true that there are links between the Earth's climate and the concentration of chlorofluorocarbons in the atmosphere, but carbon dioxide will be the hard diplomatic nut to crack because it is the by-product of most economic activity. So why not start on that now, for fear of using up the goodwill of the world's environment ministers?

## "Endymion"

Readers persuaded by last week's reference to Wordsworth's "Endymion" that *Nature's* name must have derived from Keats should know that the reference is to one of Wordsworth's "Miscellaneous sonnets", all of which are untitled. □

# London ozone meeting wins some hearts

- 20 new signatories for Montreal
- India and China demur

## London

The London conference on the ozone layer had preached successfully to the converted when it finished on Tuesday this week, but both China and India left muttering that they would seek amendments to the Montreal Protocol at the meeting planned for Helsinki in May.

The conference, organized by the British government and attended by 124 others, was largely an exercise in persuasion. By the time it opened, the European Community, followed by the United States and Canada, had all agreed completely to phase out by 2000 the production and consumption of the materials covered by the Montreal Protocol.

The protocol requires a 50 per cent reduction by 2000 of 5 named chlorofluorocarbons (CFCs) and 3 halons. The European Community seems to have surprised itself by its members' willingness to bring forward a complete ban. They may now seek to amend the protocol in May.

The fear now among the industrialized nations is that their attempts to halt the depletion of the ozone layer will be negated by the potentially huge increases of emissions of the restricted materials from developing countries.

Mrs Thatcher, the British Prime Minister, put the point directly when she told the conference that the actions of the developed world alone will not be enough, and said that the consequences of ozone depletion do not affect only those whose products do the most damage.

President Daniel Arap Moi of Kenya, the chairman of the opening session, put forward the poorer countries' view by saying that many developing countries are only now embarking on the large-scale use of refrigeration, air-conditioning, plastics and electronic manufacturing in which the ozone-depleting chemicals are used. He said that industrialized countries would have to help them leap-frog the CFC era, and that industrialized countries should be prepared to make sacrifices commensurate with those now being asked of developing countries.

Even so, 20 countries announced at the conference that they would sign the protocol, bringing the number of signatories to 59. (Only 32 have so far ratified the protocol.) It is especially significant that Brazil, one of the most rapidly industrializing of the developing countries, should have agreed to sign, which directed the attention of the conference to India and China.

Z.R. Ansari, the Indian Minister for

Environment and Forests, and Lui Ming Pu, of the Chinese commission for environmental protection, both spoke in similar terms, complaining that the restrictions placed on developing countries by the protocol are too severe, and that the provisions for technical assistance should be spelled out in more detail.

Both representatives said their governments would be willing to sign the protocol if a fund were set up by industrialized countries to support research and development leading to alternative materials, and if the outcome were transferred to developing countries free of charge. But China will be represented at Helsinki to explore amendments of the existing protocol.

Others may then be reluctant to see the protocol along the lines now accepted by the European Community, the United States and Canada. Thus Vladimir Zakharov, deputy chairman of the state committee of the Soviet Union for hydrometeorology, said that more scientific evidence is needed before tighter restrictions are imposed.

But the European Community is already pressing to go further than the position agreed at its meeting last week, on the eve of the London conference. Environment Commissioner Ripa di Meana said that the aim should be a complete phase-out of CFCs and halons by 1996-97. Mr Nicholas Ridley, Britain's Secretary of State for the Environment, and Klaus Toepfer, his counterpart from West Germany, said they would support such a move. US Senator Albert Gore (Democrat, Tennessee) is also in favour of a tighter 4-5 year deadline.

Robert Watson, from the US National Aeronautics and Space administration, would go further. He told the conference that when the protocol is amended, not only should the deadline be tighter but that there should also be a ban on the emission of carbon tetrachloride as well as tight restrictions on methyl chloroform.

Representatives of industry are rattled by the prospect of repeated changes of legislation which affects their business. They say that while a phase-out by 2000 is feasible, further changes could impede the development of substitutes. Archie Dunham, a vice-president of E.I. du Pont de Nemours, said that without a secure knowledge that there will be a long-term market for potential substitutes, the phase-out of chemicals now in use will be delayed.

Christine McGourty

## CFC SUBSTITUTES

### Electronics industry in front line

#### Santa Clara, California

REPRESENTATIVES of the US electronics industry, faced with the prospect of an enforced phase-out of the CFC solvents used as cleaning and drying agents in manufacturing, met CFC producers last week in Santa Clara to try to speed up the search for alternatives.

While the use of CFCs in aerosol propellants and refrigeration has been given most attention in the assessments of the threat to stratospheric ozone, attention has only more recently focused on CFC-113, a solvent used in more than 100 specialized applications in electronics manufacturing. Not only has the electronics industry started late in the race to develop alternatives, but it is unlikely that one or even several substitutes will be adequate replacements in all applications.

The electronics industry seems nevertheless reconciled to a complete ban on CFC use by 2000. The US administration, which has been reluctant to call for measures going beyond the 50 per cent reduction specified in the 1987 Montreal Protocol, lined up with the European Community at the weekend in recommending a complete phase-out by 2000 (see left).

If the recommendation is not formally endorsed by the signatories of the protocol at the meeting planned for Helsinki in May, the US Congress could decide to act unilaterally on behalf of the United States.

While most industry representatives at the conference here accepted that a CFC phase-out is necessary to minimize ozone destruction, they fear that the regulatory axe will fall before they have safe economical substitutes in place. Hasty decisions on replacements, they say, could themselves create problems in the future, should the new chemicals prove toxic or otherwise hazardous, for example.

A unilateral ban would in any case be no more than a 'band-aid', according to David Chittick, vice-president for environment and safety engineering at AT&T, who said that such a ban would penalize US companies economically while doing little to save the ozone.

Some CFC suppliers with CFC-113 replacements to describe at the meeting complained that the electronics industry is reluctant to consider alternatives that will require costly changes in equipment and manufacturing processes, and seems rather to be waiting for a 'magic liquid' that can be used in a fashion identical with that of existing products.

But this week both du Pont de Nemours and the British company ICI announced the development of alternatives to CFC-113, and AT&T says it has developed a cleaning method that eliminates the need for CFC-113.

Marcia Barinaga



## Cattle disease set for cure

### London

BRITISH veterinarians were congratulating themselves last week that a disease of cattle first recognized in 1986 could have disappeared again by 1993. That is the prediction of a working party under Sir Richard Southwood set up last year to study Bovine Spongiform Encephalopathy (BSE), a fatal neurological disease of cattle. But the prediction hangs on the enforcement of last year's regulations to prevent the feeding of cattle with feed-stuffs derived from sheep infected with scrapie.

BSE is an exclusively British disease, now affecting 1 in 1,000 cattle each year. Reports of similar cases from the United States are said not to be conclusive. The working party says the cause of BSE is the agent responsible for scrapie, the spongiform encephalopathy of sheep. It is confident that the agent is not transmitted between cattle, either *in utero* or between adults.

On the risk that BSE may be transmitted to people, when it would presumably show up as a condition similar to Jacob-Creutzfeldt disease, the working party says that it has warned the Committee on the Safety of Medicines of the need for scrutiny of the sources of bovine medicinal products. Otherwise, the group asks that meat or milk from infected cattle should be kept out of the human food chain (already required by last year's legislation).

On the principle that bovine lymphatic tissues, followed by brain and nervous

tissue, are likely to be first affected by the BSE agent, the working party considered, only to reject, the proposal that food for human consumption containing such material should be so labelled. But the government has accepted the working party's proposal that it should legislate to keep these materials out of baby food.

Several factors are thought to have been responsible for the emergence of BSE in Britain, including a significant increase of the sheep population in the 1980s, a possible increase in the incidence of scrapie, the greater use of sheep heads in rendering (the process of extracting fat and protein from animal wastes) and new rendering processes, which may be less effective at inactivating the scrapie agent.

Not everybody accepts all the working party's conclusions. Dr James Hope, director of the neuropathogenesis unit at Edinburgh, jointly supported by the Agricultural and Food Research Council and the Medical Research Council, says that *in utero* transmission is likely, although there is as yet no proof. Hope would also label food products containing bovine lymphatic and nervous tissues so as to give consumers an informed choice.

But Hope welcomes the working party's advocacy of a research programme further to study the source and transmission of the disease. The government announced last week that a committee under Dr David Tyrrell will advise on research in progress and on priorities for the future.

Christine McGourty

### JAPANESE JOBS

## Students prefer finance and pay

### Tokyo

THE glitter and excitement of the financial markets is proving just as strong a lure to bright Japanese graduates as to their counterparts in the West. A survey\*, released last week by the National Institute of Science and Technology Policy (NISTEP) shows a sudden downturn in science and engineering graduates entering manufacturing industry and a surge in those heading for Kabuto-cho, Tokyo's Wall Street.

At the same time, figures from the Ministry of Education, Culture and Science (MESC) show that engineering students are increasingly choosing salaried jobs rather than further education in the universities. As a result, graduate engineering schools are coming to be dominated by foreign students, principally from China and Taiwan, just as in the United States (see *Nature* 336, 102; 10 November 1988). In some electrical engineering departments, there are already more foreign doctoral students than there are Japanese.

The NISTEP analysis shows that science and engineering graduates entering manufacturing industry increased steadily over the two decades until 1987. Then came a sudden drop of 5–6 per cent, accompanied by a surge in numbers entering finance, insurance and property companies. A follow-up survey of ten representative universities charts the same phenomenon: engineering graduates entering manufacturing industry fell from 76 per cent in 1986 to 66 per cent in 1988, while the percentage going into finance and insurance more than tripled.

NISTEP is not ready to say whether the shift marks a long-term change or a temporary oscillation. But it is enough to worry industrial companies which view Japan's technological edge as all important. Countermeasures, including more flexible working hours and better work conditions, are being planned. But pay levels, which are really what matters, cannot be increased without radical change to company seniority systems.

The influx of foreign students is both an

## French telescope seeks Indian mountain

### New Delhi

FRANCE has proposed moving one of its optical telescopes to India as the first step in a collaborative research programme in astronomy. The proposal was made by Professor Jean Audouze of the Institut d'Astrophysique de Paris at a recent conference in New Delhi of Indian and French astronomers.

The 10-year-old 1.5-metre Schmidt telescope is now at the Centre d'Etudes et de Recherches Geodynamiques and Astronomiques at Grasse, on the French Riviera. French astronomers have been looking for a new location for the telescope because Grasse, at an altitude of 1,200 metres, suffers from light-pollution from nearby Nice. Attention has focused on the Himalayan region.

Dr Jayant Narlikar, an astrophysicist at the Tata Institute of Fundamental Research in Bombay, says that relocating the telescope will be a boon for astronomers in India, where there is no such instrument.

As well as better viewing conditions, putative sites in India offer larger coverage of the southern sky. Surveys have indicated that Drusthal, near Nainital in Uttar Pradesh state at a height of 2,200 metres, will be acceptable. It has 250 clear nights a year compared with 100 at Grasse. It has been proposed that India and France will divide the cost of resiting the telescope, which will be used by both countries.

A second proposal is for a joint experiment between the University of Pune in India and the Meudon Observatory in France for detecting gravity waves using very-large-baseline interferometry with one antenna in France and another in India.

K.S. Jayaraman

immediate stimulus and a longer-term worry. About 25,000 foreigners are now studying in Japan (expected to increase to 100,000 by 2000), almost three-quarters of them coming equally from China, Taiwan and Korea. One-fifth study engineering, mostly at the postgraduate level. By that stage many Japanese engineering students have gone to industrial companies where facilities are better and the pay is higher.

Engineering professors praise the ability and motivation of their foreign students but worry what will happen when they go home. Unlike the United States, where universities happily accept foreigners as permanent staff, Japanese universities remain relatively closed. Without change, engineering faculties may face a decline in their senior research staff just as Japan's high technology is meant to be conquering the 21st century.

Alun Anderson

\* Preliminary Survey of Employment Trends for Science and Engineering Graduates soon available in English from NISTEP, 1-11-39 Nagata-cho, Chiyoda-ku, Tokyo 100, Japan.

## AMAZON FORESTS

# Japan and Brazil team up

## Washington & Tokyo

JAPAN'S hunger for tropical timber, of which it is the world's largest consumer, has drawn its attention to the Amazon rain forests and onto a collision course with international organizations concerned about the shrinkage of the tropical forests.

Japanese imports of tropical timber from Brazil began last November and Japan's greatly increased overseas aid (just under \$10,000 million this year) may soon help to finance a road linking the upper reaches of the Amazon Basin with the Pacific coast of Peru, opening up vast tracts of the forest to exploitation. But environmentalists and US legislators concerned about the possible adverse

fically mention the Acre road.

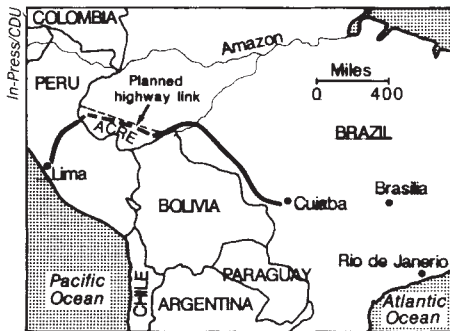
Japanese Foreign Ministry officials say no agreement to finance the road has been made. But Japan is ready to make huge loans to Brazil and, according to the Washington-based Environmental Defense Fund, the terms of earlier loans to Brazil have often been breached.

Kasten's legislation is sure to draw attention to Japan's recent moves into the Amazon. After a quiet beginning last November, Japanese imports from Brazil soon caused great embarrassment for the two countries when disclosed a few weeks later at a meeting of the International Tropical Timber Organization (ITTO) in Yokohama. In September, Brazil's president Jose Sarney announced an export ban on raw logs so as to develop an indigenous timber-processing industry and to slow the destruction of Amazonian

forests. But Japan circumvented the ban (with Brazil's approval) by taking timber from areas cleared for a hydroelectric project financed with overseas aid.

Japan now gets most of its timber imports from South-East Asia. But supplies are beginning to dry up. Indonesia and the Philippines have banned exports of raw logs and have begun processing their own timber. In response, Japanese trading companies switched attention to Malaysia. But Malaysia's tropical forests are rapidly being depleted, and Japan is now turning to Brazil, and to the Amazon.

Now the world's largest overseas aid donor, Japan plans to channel thousands of millions of dollars into South America, in particular Brazil. World Bank and IADB loans to developing countries are already under severe scrutiny because many giant projects have had unforeseen environmental side-effects. Japan can expect the same tough criticism of its loan plans. **David Swinbanks & Alun Anderson**



A new link to Pacific ports could cut time needed to ship out timber from the Amazon.

effects of deforestation on the world's climate plan to fight the proposal.

The new road, backed by Flaviano Melo, governor of the Amazonian state of Acre, would link up with the trans-Amazon highway which connects ports on Brazil's Atlantic coast with Acre. But that highway, built with funds from the World Bank and the Inter-American Development Bank (IADB), brought destruction to Brazil's rainforests as more than a million migrant peasants moved up the road, burning and clearing adjacent forest for temporary subsistence farms. Subsequent IADB loan disbursements for road improvements in Acre were blocked after opposition in the US Senate. So Melo has turned to Japan for money.

Senator Robert Kasten (Republican, Wisconsin), who was instrumental in bringing World Bank and IADB support to a halt, announced at a Senate Commerce Committee hearing on global climate change last week that he has introduced legislation calling on the US Secretary of State to oppose Japan's "reported" plans to finance the new road with overseas aid. The matter was taken a step further when US President George Bush, in Tokyo for the funeral of the late Emperor Showa, reminded Japan's Prime Minister Noboru Takeshita of concern over global environmental problems and deforestation, although he did not speci-

## RESEARCH ANIMALS

# Wild chimpanzees endangered

## Washington

WILD chimpanzees, previously classified as 'threatened', have now been reclassified as an 'endangered' species by the US Fish and Wildlife Service. As a result, it will now ordinarily be illegal to import into the United States chimpanzees or their young which have been caught in the wild. But captive chimpanzees now in US zoos and biomedical research facilities will remain in the 'threatened' category, which entails fewer restrictions on movement.

Because of their likeness to human beings, chimpanzees have been valuable animals for biomedical research, raising serious concern about the survival of wild chimpanzee populations.

The new rules were prompted by a petition filed more than a year ago by the Humane Society of the United States, the World Wildlife Fund and the Jane Goodall Institute. Flooded by comments supporting the proposal, the Fish and Wildlife Service announced its intention to reclassify the chimpanzee late last year (see *Nature* 336, 511; 8 December 1988).

One significant subtlety in the rules now published is that, while captive chimpanzees in the United States will remain classified as 'threatened', captive chimpanzees in Africa will be redesignated as 'endangered'. Conservationists hope this will prevent poachers from selling newly captured animals.

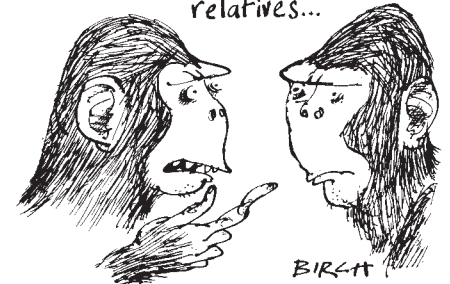
The new classification is not universally acclaimed. The US National Institutes of Health (NIH) have opposed it, arguing that there should be a systematic survey of wild chimpanzees before the import restrictions are tightened. But conservationists have denounced this pro-

posal as a delaying tactic to keep open the door by which wild chimpanzees could reach research laboratories.

With stricter rules now in place, Curtis Bolin says that the World Wildlife Fund will test NIH's sincerity by offering to join it in a new study of wild chimpanzees in Africa. Deputy director William Raub says that NIH may be interested in surveying several other primate species as well.

Under the new rules, all major facilities in the United States housing captive chimpanzees will have to account each

*You know what I really hate?  
their claim to be our closest living  
relatives...*



year for the numbers of their chimpanzees as well as their health and breeding status.

Conservationists' suspicions that biomedical researchers have been subverting efforts to protect the chimpanzee have been rife since Robert Gallo told an international conference several years ago that he would favour importing wild chimpanzees for use in AIDS research. Gallo is now seeking to mend fences: a statement by him that AIDS research no longer needs chimpanzees from the wild is expected to appear soon in the popular magazine *The New Yorker*. **Carol Ezzell**



## British make £11 million claim

London

THE Clinical Research Centre (CRC) of the Medical Research Council (MRC) has been chosen as the hub of Britain's part in mapping and sequencing the human genome. CRC will house a resource centre responsible for assembling and distributing databases and DNA libraries as well as for substantial technical developments, both in computing and sequencing.

The planners hope their decision will also attract the administrative centre of the Japanese Human Frontiers Science Programme and the European office of

the Human Genome Organization (HUGO) to the same site at Northwick Park in north London.

MRC's decision follows the British government's allocation of extra research funds for the next three years, of which £11 million has been earmarked for the genome project. While this is small compared with US spending (\$27.6 million this year and \$100 million next), it is as much as the European Community plans to spend (see accompanying story).

Sydney Brenner, the chief architect of the plan, says that the arrangement should

give Britain a voice in international discussions. "We will now have trade goods", he says. Brenner's own laboratory at Cambridge will focus on the isolation and characterization of new cDNAs, while the resource centre will concentrate on building a library of accessible DNA clones.

The choice of CRC will be controversial. CRC is due to merge with the Royal Postgraduate Medical School at Hammersmith and is not particularly strong in human genome work. But MRC plans to redeploy resources at CRC into human genetics, while there is ample space both for MRC's own centre and potential tenants.

MRC also plans to guide the overall direction of British human genome research by a coordinating committee and a scientific advisory board, whose priorities will influence which projects it finances. The board will be much like that which, since 1987, has coordinated the human genome interests of MRC and the Imperial Cancer Research Fund (ICRF).

ICRF, which has been keen to bid to provide the MRC resource centre with a DNA probe bank, has developed a particular interest in the European Community project and is also a partner in the EUREKA project to automate DNA cloning and sequencing. **Peter Newmark**

### GENOME MAPPING

## Europe's plans turn towards talk

Munich

THE European Community's embattled (and now renamed) Human Genome Analysis proposal has been much changed by a debate last month in the European Parliament. The proposal calls for an in-

vestment of ECU 15 million (about £10 million) over three years for mapping and sequencing the human genome.

The parliament accepted most of the changes suggested by its committee on energy, research and technology led by representative Benedikt Härlin of the West German Green Party, who was a vocal critic of the original programme. Among the 38 amendments now adopted are a ban on somatic gene therapy within the framework of the programme; a study on the history of eugenics; and the dropping of most of the medical and industrial justifications in the text of the original programme.

If the amendments are enacted, much of the money for the programme will be diverted from the support of basic science and computer studies to the support of public debate over the ethics of the programme. The main purpose of the changes, says Härlin, is to prevent "the money from simply being thrown into technology". Instead, there should be "political control" over the programme from the beginning.

An official at Brussels says that the commission would accept most of the proposed amendments in some form, especially those strengthening the rights of persons to protect genetic information about themselves. But there will probably be some dispute in the member states over some of the other amendments — they might not go far enough to satisfy West Germany, but they might go too far for the British and French.

The proposal must now be discussed by the European council of research ministers before being sent back to the parliament for the next round of debate. Once it has the approval of a majority of the 12 council members and the parliament, the proposal can finally be enacted by the commission. The intended starting date, 1 January 1989, will then be even more distant. **Steven Dickman**

### US PROJECT

## Bio-information centre

Washington

DAVID J. Lipman has been appointed director of the new Biotechnology Information Center set up within the National Library of Medicine, part of the US National Institutes of Health (NIH). The management of NIH is known to be dismayed that the centre has been established separately from the office established to coordinate efforts to map and sequence the human genome.

The centre was established last year by legislation authorizing the human genome project, and will develop software and database systems for handling the large amount of data emerging from the project. Congress appropriated \$8 million for the centre for 1989, some of which will go towards partial support for the GenBank and Protein Information Resource databases. Twelve bio-informatics specialists will work at the centre, but James Ostell, the developer of the Pustell DNA sequence manipulation software, is the only one hired so far.

Lipman developed software for searching sequence databases while working elsewhere at NIH, and has served on the advisory board for GenBank. He will be coordinating the centre's programmes with the information advisory subcommittee of the NIH Human Genome Office, composed of David Botstein from Genentech, Mark Pearson from DuPont, Jaime Carbonell from Carnegie-Mellon and George Cahill from Howard Hughes Medical Institute.

Carol Ezzell

### NEWS IN BRIEF

## Research! America

Washington

FORMER US Senator Lowell Weicker will head a newly formed alliance of health-related organizations designed to drum up public support for biomedical research. The alliance, dubbed *Research! America*, will not lobby Congress for support, but will instead conduct a campaign to persuade people that biomedical research is worth the money being spent on it, that it needs more money to continue to thrive and that becoming a researcher is a worthwhile career goal. Weicker believes the real remedy for runaway health care costs will be research for new therapies, not a band-aid approach to health care financing.

Weicker's presence at the head of *Research! America* should help it get off the ground. As a Senator, he was a vocal proponent of research, and a strong supporter of the National Institutes of Health. The alliance expects to raise about \$2 million per year. **J.P.**

## Military review

Paris

THE French minister of defence, Jean-Pierre Chevènement, has ordered a six-month review of defence research. The study will be carried out by Jean-François Delpech, research director at the CNRS (Centre National de la Recherche Scientifique). At FF30,000 million (\$4,840 million), the defence budget is about one-third of French research spending. **P.C.**

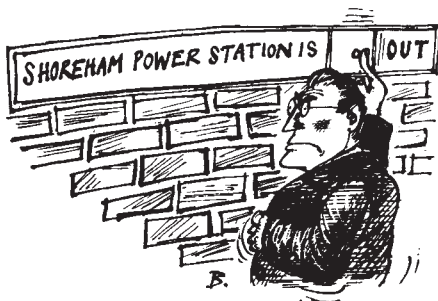
# Risen phoenix bites dust again

**Boston**

NEW YORK's Shoreham nuclear plant seems again to be heading for the scrap-heap. In a major surprise for all parties involved, New York Governor Mario Cuomo last week signed an agreement with the Long Island Lighting Company (LILCO) that would permanently close and dismantle the Shoreham nuclear power plant.

But even as the ink was drying on the new agreement, the Nuclear Regulatory Commission (NRC) removed the last obstacle to issuing an operating licence for the Shoreham plant. While most observers believe that the shutdown agreement will hold even if Shoreham is now licensed, the timing of recent events can only increase the confusion about its future.

Announcing the new agreement last week, Cuomo said he expected the new plan to "end this matter once and for all". But the tangled history of negotiations to close Shoreham and the curious timing of



the NRC ruling has made local residents, investors and regulators doubtful that the end has really arrived.

A similar plan to shut the plant last spring fell through unexpectedly when the state legislature failed to endorse it. Furthermore, the new agreement comes just as the completed but never-licensed

EATING FOR HEALTH

## Diet in the United States

**Washington**

IF THE National Research Council's report on diet and health is not the final word on the subject, it is certainly the weightiest. The report, released last week, weighs nearly 6.5 pounds in a paper bound version. In addition to providing a thorough assessment of the health risks associated with various foods, there are dietary recommendations with a familiar ring; reduce total fat intake to 30 per cent or less of calories, eat five or more servings per day of fruits and vegetables, moderate protein consumption, limit salt consumption to 6 g per day or less. Implementing these and other recommendations is left to another panel now preparing a report due out in a year.

Joseph Palca

nuclear plant had begun to show signs of life (see *Nature* 338, 3; 2 March 1989).

In the shutdown agreement, the state of New York will purchase the \$5,500-million nuclear plant from the utility for \$1 and will also assume responsibility for decommissioning and dismantling it. The new agreement is almost identical with last year's plan, except that the closure does not now require the approval of the state legislature. Unlike the previous plan, the new agreement will also allow the State Public Service Commission to decide what are appropriate rate increases for LILCO's electricity consumers. Last year's abortive agreement included an explicit schedule of increases.

By signing the agreement, Cuomo has given point to his long-standing opposition to Shoreham. For this decisive action he has been applauded by many of those involved in the controversy, including utility representatives who have said repeatedly that prolonged uncertainty about the future of the plant would be the most costly and damaging outcome. But in his latest move, the governor has also left himself open to criticism and unpopularity from Long Island residents, who face increased electric prices and an expected two years of electricity shortages.

The Shoreham decommissioning agreement has been signed by both Cuomo and the president of LILCO, but it still needs the approval of LILCO's board of directors and shareholders, which is likely to be given. But the effect of the ruling by the Nuclear Regulatory Commission is still unclear. NRC has disqualified New York State and Suffolk County from participating in the licensing process, effectively overriding state and local protests that the plan for evacuating the local population from the neighbourhood of the plant is inadequate. Cuomo called the ruling "absurd" and "academic", noting that the agreement to shut the plant had already been signed.

Even so, given the ups and downs of Shoreham so far, the loopholes in the closure agreement cannot be ignored. Most prominent among these is the likelihood that LILCO's outstanding application for an operating licence for Shoreham will be approved before April, when the state's Public Service Commission is required to settle the rate increases the utility will be allowed. If LILCO is not satisfied with them, it could conceivably renege on the agreement.

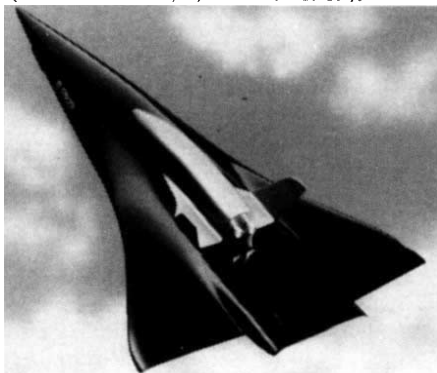
LILCO representatives nevertheless claim to be committed to the new agreement, and refuse to "speculate" about other possible developments. Yet for many observers of the controversy, Cuomo's new closure plan will be final only when a start has been made on dismantling the plant.

**Seth Shulman**

## Full speed ahead with space plane

**Munich**

WEST German Research Minister Heinz Riesenhuber announced last week that West Germany will press ahead with the development of a space plane called Sanger (see *Nature* 338, 6; 2 March 1989).



The plane, named after German aerospace pioneer Eugen Sanger (1905–1964), would consist of two reusable parts: a booster stage resembling a conventional jumbo jet would bring a smaller craft piggyback to an altitude of 30 km. The smaller plane would then fly at five times the speed of sound to its destination. Sanger could be used for passenger transport on Earth or as a space shuttle.

The Research Ministry plans to spend DM220 million on the project over four years, with the West German aerospace research agency DFVLR contributing an additional DM85 million and private industry DM2,530 million. Riesenhuber estimated that the cost of transporting a kilogram of payload into space could be reduced from the current \$8,000 to \$2,000 or even \$1,000.

Britain, France, Japan and the United States are currently working on space transporters of their own. Sanger, the British HOTOL or the French contribution might eventually be adopted by the European Space Agency as a successor to the Hermes shuttle currently being developed.

Steven Dickman

**FRANCE**

## Hazard prevention

**Paris**

MICHEL Rocard, French prime minister, has set up a commission for the prevention of industrial and technological hazards. The aim of this new body, which is composed of twelve experts from potentially hazardous disciplines, ranging from biology to mining, is to reduce public anxiety as well as to prevent accidents involving dangerous substances. The government already has a Secretary of State for the Prevention of Major Technological and Natural Hazards as well as organizations for nuclear safety and an ethics committee with a wide-ranging remit.

Peter Coles



# US plans next electronics war

## Washington

THE US electronics industry has shaken off its lethargy and taken the first strides in a race with Japan and Europe to develop high-definition television (HDTV), a technology that may revolutionize not only television but also the computer and telecommunications industries. Last week, American Telephone and Telegraph (AT&T) and Zenith Electronics Corporation, the only remaining US-owned television manufacturer, agreed to develop jointly an HDTV system. They have jointly applied to the Defense Advanced Research Projects Agency (DARPA) for support.

HDTV is also becoming a hot issue in Congress; last week saw the introduction of legislation intended to encourage the formation of a government-supported consortium to develop the new technology. Although the United States is late in joining the HDTV race, many in US industry and universities are confident they can leap-frog Europe and Japan.

HDTV, which Japan's National Broad-

casting Corporation (NHK) began developing in the early 1970s, has twice as many lines on the screen as conventional television, and thus offers pictures with greater clarity and resolution. HDTV also has applications in the movie industry, military imaging systems (whence DARPA's interest), still photography, printing, computer graphics and microscopic imaging. And supporters of HDTV talk of linking it with personal computers to create new information media.

If HDTV takes off, it will create a huge market for the semiconductor industry because of its dependence on computer chips. But present versions of HDTV receivers have drawbacks likely to deter consumers. They are heavy, bulky and expensive, while the high quality of the picture is apparent only when the screen is large. This is where US researchers see an opportunity to enter the market.

Zenith, as well as applying to DARPA for \$13 million to develop HDTV receivers in collaboration with AT&T, has also applied for \$10 million to develop a

large-screen version of its 'flat tension mask' tubes. It claims these will be cheaper than other display systems.

John Taylor of Zenith says that if the joint venture with AT&T is fully supported, the companies will be able to put a television on the market for between \$2,500 and \$3,000 in 1992-93. The price of the first Japanese models is about \$12,000.

William Glenn of the New York Institute of Technology also sees opportunities in the development of flat display screens that can, for example, be hung on a wall. Meanwhile, researchers at the Media Laboratory of the Massachusetts Institute of Technology (MIT) are developing an 'open architecture' computerized HDTV that can be programmed to receive and display signals based on any of the transmission standards being considered in the United States, Europe and Japan.

Budget pressures may limit what can be done. DARPA's programme, announced by the Department of Defense only in December, has only \$30 million to spend this year, and cannot support all the 80 projects received.

Ironically, although the programme was introduced to spur US initiatives, many of the applicants are foreign-owned companies based in the United States. One is Japan's Sony Corporation, which is asking for the full \$30 million to develop high-resolution displays and image-processing technology based on the NHK standard.

One possibility is that DARPA could use the funds to establish another government-supported consortium on the lines of SEMATECH, which is given DARPA funds to develop next-generation computer chips. But many in the Pentagon are opposed to DARPA's ventures into the civilian sector, while even the agency's supporters in Congress see the need for new machinery, separate from the defence establishment, for promoting civilian research and development of strategic technology.

Such is the objective of the HDTV bill introduced in the House of Representatives last week by Congressman Don Ritter (Republican, Pennsylvania). The bill would modify antitrust laws to allow US companies to enter joint ventures in the development and manufacture of HDTV. The bill would also offer tax incentives to companies developing HDTV products and would also encourage the administration to provide cost-sharing finance for a consortium of domestic companies.

But even if the United States manages to leap-frog Japan and Europe in the development of HDTV, will it be able to hold the market? Japanese companies have shown themselves to be remarkably adept at penetrating the US market with high-quality low-priced products even when they come from behind. **David Swinbanks**

## TECHNOLOGICAL RIVALRY

# US lead perceived and envied

## Tokyo

US INDUSTRY'S alarm about Japan's lead in the race to develop high-definition television (HDTV) is unlikely to evoke much sympathy from Japanese companies, which are themselves concerned at the number of fields of high technology in which the United States is dominant.

Recent surveys and the latest white paper (policy document) from the Ministry of International Trade and Industry (MITI) have rather stiffened the determination of Japanese companies to undertake more basic research to help bridge the perceived gap between Japan and the United States.

Surveys serialized last week in Japan's leading economic and industrial newspapers (*Nihon Keizai Shinbun* and *Nihon Sangyo Shinbun*) show that Japanese companies believe that in basic research they are ahead of the United States only in the general field of optoelectronics. They believe there is a large gap between Japan and the United States in the life sciences and lesser gaps in the development of computers and new materials.

In an analysis of 47 sub-areas of high technology, Japan believes it has a clear lead over the United States only in high-powered magnets and intelligent robots. In everything else, from high-temperature superconductors through biotechnology to software, the United States is believed to lead. By the year 2000, the United States is seen as still dominant but with

Japan adding leadership in fine ceramics, superconducting devices, carbon materials, ultra-large-scale integrated circuits, new bio-materials, disaster prediction and environmental management.

The really bad news is for Europe, which Japanese companies believe to be well behind the United States in all areas of high-technology research. In 12 of 13 areas of biotechnology, Europe comes ahead of or equal to Japan. But the lead in these areas is expected to vanish by 2000.

The *Nikkei* survey complements MITI's most recent white paper. A detailed analysis of the front-line of technological research puts Japan behind the United States not only in the basic research areas explored in the *Nikkei* survey, but also in areas requiring very-large-scale integration. These include space rockets and satellites, nuclear reactors and civil aircraft.

The MITI white paper seeks success by long-term planning for the generic technology of the next century, by which time the US panic over HDTV may appear misplaced. Although Japan may be in the lead (the first Hi-vision satellite broadcasts are planned for next year and a MITI-led consortium is already developing large flat display screens), HDTV is regarded less as a new and dominating generic technology than as an opportunity for world-wide suppliers of everything from satellites to microchips.

**Alun Anderson**

## French Assembly to lay down the law

### Paris

A DISCUSSION document to be presented to the French National Assembly in the spring could, if accepted, lead to significant changes in civil rights laws relating to organ transplants, biomedical research, *in vitro* fertilization and personal privacy. The draft law, entitled "Life Sciences and Human Rights", has been produced by a working party set up only last December by Michel Rocard, the prime minister.

The proposal would add to existing legislation on civil rights a new definition providing for the first time a right to the "integrity of the human body". The draft article concerned states that nobody may violate a person's right to the integrity of the human body without his or her consent, unless there is a therapeutic emergency in which consent may not be obtainable.

While the same draft article states that organs or human by-products cannot be subject to ownership claims, or to remuneration (in the case of donors), Article 4 of the proposed act endorses the 1976 'Cavallier Law' of presumed consent. That allows organs to be removed for transplant from any person judged clinically to be dead, unless the person had expressed wishes to the contrary. This article would also make it illegal for minors to donate organs except to siblings, in which case the opinions of three physicians must be given.

The draft proposes amendments to the existing laws on *in vitro* fertilization (IVF). It prohibits payment for donated sperm or ova and outlaws the practice of surrogacy, stating that an (infertile) woman can have no claim of maternity if a third party should bear a child ostensibly on her behalf. But the husband or concubine of the surrogate may refuse to recognize the child.

It is also proposed to regulate the conservation of supernumerary embryos arising in the practice of IVF. Unwanted embryos would usually be allowed to develop only for five days (14 days with the approval of the national ethics committee) and may not be used for eugenics experiments or in ways that "violate the integrity of the human race".

Frozen embryos would have to be destroyed after five years or if the putative parents should cease to be a couple (through death, divorce or separation). But such embryos could be donated by these parents to another couple.

The proposed law would also redefine the constitution, independence and impartiality of the national committee on ethics. The last of the six draft articles affirms the right of access by epidemiologists to medical files, but also gives patients the right to be informed when the use of personal information is sought and to oppose it if they choose. Peter Coles

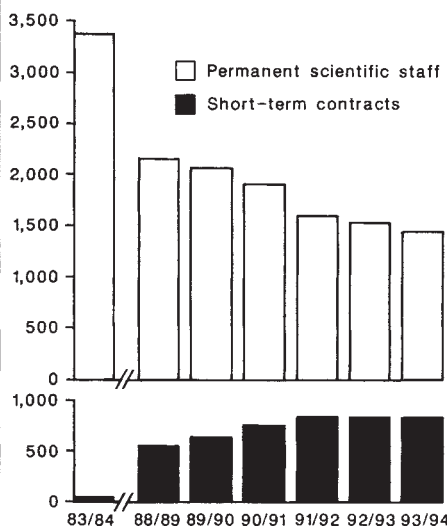
## More without job tenure

### London

Two British research councils are planning still further to increase the proportions of their researchers on short-term employment contracts. These trends, acknowledged by the Natural Environment Research Council (NERC) and the Agriculture and Food Research Council (AFRC), arise from plans to channel more resources towards universities. But the councils' managements are conscious of the long-term dangers.

Professor John Knill, chairman of NERC, said last week that the open market for scientific manpower in the European Community from 1992 could mean an influx to Britain of foreign scientists seeking research training and an exodus of highly trained researchers looking for permanent posts abroad.

Short-term appointments have increased dramatically in recent years. At AFRC, the number on short-term appointments in the council's own institutes has grown from 30 in 1983 to almost 600 now, and is planned to be 850 in 1994. But the



AFRC staff complement — permanent appointments are dwindling.

number of permanent posts will decrease by 700 in the same 5 years and 500 supporting posts will also go.

The trend at NERC is in the same direction, if less swift. Total manpower has decreased by 18 per cent in the past five years (from 3,300 to 2,700), but short-term appointments (both at the council's own institutes and on its university payroll) have increased from 200 to 350. Both trends will continue. Despite this year's budget increase, NERC expects to have to reduce its staff by 100 a year indefinitely.

Professor Bill Stewart, deputy chairman of AFRC, says that short-term employment contracts will give his council the flexibility to increase the university proportion of its research spending to 30

per cent in the next five years. To help meet industrial demands for skilled manpower, he says AFRC is also planning an extra 200 studentships (for graduate students), mostly in molecular sciences, as well as a number of research fellowships and professorships.

By contrast, NERC's hopes of increasing support for university research seem to rest on the outcome of a request for an extra £7 million in its budget for next year, which would be spent on research fellowships and grants. Following the example of the Medical Research Council (see *Nature* 338, 7; 2 March 1989), AFRC also plans to increase the value of its studentships by £500 a year from next September. This is an interim measure: all the research councils are now hoping to put into effect an increase of £1,000 a year in the academic year 1990-91.

Christine McGourty

### INDIA'S BUDGET

## Modest science boost

### New Delhi

SCIENCE does modestly well in the Indian government's election-year budget presented to parliament last week. Seven scientific departments between them will have \$1,460 million to spend in the financial year beginning on 1 April. That amounts to 2.7 per cent of the total Indian budget and \$150 million more than last year's support for science and technology.

Predictably for an election year, the budget is orientated towards the poor and unemployed, with less emphasis on expansion of the scientific empire. As in previous years, the Department of Atomic Energy (DAE) gets the lion's share (52.5 per cent) of what there is. The budget for the Department of Space (DoS) has been marginally reduced. On the other hand, 25 per cent more than last year will be spent on biotechnology, chiefly to continue projects for producing biological insecticides against mosquitoes, interferon through genetic engineering and for growing bamboo by tissue culture.

For the first time, India plans to cut its defence expenditure — to \$8,700 million, a drop of \$130 million from last year.

The Indian Institute of Science in Bangalore will get \$19 million for a new supercomputer "for enhancing capacity in virus crystallography and bioengineering". A token provision is also made for the \$160 million DoS project to develop a cryogenic rocket engine with 12 tonnes of thrust. And the DAE has been granted \$140 million for a new uranium oxide fuel plant of 50-tonne annual capacity. Notably absent from the budget are funds for projects in materials science.

K.S. Jayaraman



# Hackers revealed as spies

## Munich

AN unprecedented blend of computer crime and espionage came to light last week when West German police arrested eight men on charges of providing passwords and other data from Western military computer systems to Soviet-bloc intelligence agents.

The 'hackers', all between 25 and 35 years old, were arrested after two of them had confessed and after police had searched a dozen houses and flats in Hanover, Hamburg and West Berlin. The group included the West German accused of breaking into hundreds of US computer systems last year apparently seeking classified information (see *Nature* 333, 105; 1988). The man, named as Marcus Hess, was not prosecuted in West Germany because of lack of evidence. The US Federal Bureau of Investigation also produced no arrests.

West German authorities refused to reveal the identity of the systems broken into by the hackers, but a report on the West German television network Norddeutsche Rundfunk said that they included a US Department of Defense database, a computer of the National Aeronautics and Space Administration (NASA) and a computer at the Los Alamos National Laboratory in the United States. The television network reported that computers at European

research institutes and defence-related companies were also entered.

Astronomer Clifford Stoll first spotted the attempts by West German hackers to obtain access to secret information while at the Lawrence Berkeley Laboratory in California. Stoll, now at the the Harvard Smithsonian Center for Astrophysics, said last week that the intruders obtained "sensitive but not classified information" from the US Department of Defense's Optimus database and elsewhere. He says that the hackers learned of US nuclear and biological warfare capabilities and also discovered the passwords of many computer systems.

Police confiscated computers and floppy disks, but released seven of the eight suspects because there was no fear of them fleeing or destroying evidence. The suspects face a maximum sentence of five years in prison.

West German federal prosecutor Alexander Prechtel of Karlsruhe has sought to play down the seriousness of the offence for Western security. But he said the authorities "did not have all the facts" about the penetration of the computer systems. The US Department of Defense has not made public any information about breaches of security in the incident.

West German Interior Minister Friedrich Zimmermann said the case was an "important success" for West German security forces. He said measures were being taken at military and industrial computers to prevent further intrusions. But Prechtel fears that hackers will continue to develop their techniques faster than computer systems managers can determine new ways to stop them.

While the hackers can be prosecuted under West German espionage laws, there are no statutes explicitly covering unauthorized access of computer systems in West Germany. This hindered the prosecutors' case against Hess in 1988.

Steffen Wernery of the Chaos Computer Club in Hamburg, which has not been implicated in the spy ring, said that a West German intelligence agency (*Bundesverfassungsschutz*) had approached him and other members of the club to try to win their support. In an interview for the radio network Deutschlandfunk, Wernery criticized the behaviour of the agency, saying that its "crass tactics" would scare potential informants away.

Stoll praised some hackers, including West Germans, for their contributions to knowledge about computer systems. But he criticized the unethical behaviour of others. "They don't realize that they're destroying their own playground", he said. "They cause immense damage to thousands of network computer users" by creating a fear of linking up to international systems.

Steven Dickman

# Are the good times here again?

## Sydney

AUSTRALIAN science has suddenly become popular in political circles. The government is providing an immediate infusion of funds for research, and has vowed to develop a major policy statement on science and technology.

The newly favourable wind appears to stem from the response of the increasingly vocal research community to a review by an interdepartmental committee (IDC) of the resource and policy needs of Australian science. The IDC presented its report to the cabinet last December. Although it has not been made public, it apparently offered little respite from the spiral of declining support for research (see *Nature* 337, 6; 5 January 1989).

Researchers' protests have been effective; Mr Barry Jones, Minister for Science, the was instructed to make a more favourable submission to the cabinet.

One immediate result is that the government has agreed to allow the Commonwealth Scientific and Industrial Research Organization (CSIRO), the Australian Nuclear Science and Technology Organization (ANSTO) and the Australian Institute of Marine Science (AIMS) to retain all their external earnings. This will result in an increase in CSIRO's budget of A\$2.1 million for 1988-89, rising to A\$6.6 million and A\$11.6 million in the two following years.

The government will also provide A\$8.5 million for 1988-89 to improve equipment at government research facilities. The bulk (A\$5 million) will go to CSIRO; the rest will be divided among AIMS, ANSTO and the Defence Science and Technology Organisation.

As suggested by the IDC review, the government will also establish a science coordinating committee to oversee scientific issues spanning a number of government departments. This committee will be chaired by a chief science adviser reporting directly to Jones.

In April, the government is planning to release its major statement incorporating the IDC report, the Smith report into university research requirements, the McKinnon report into marine science and a review of agricultural research.

Responsibility for the new science policy will rest with Jones, for whom the new developments represent a personal resurgence. Jones's ministerial influence declined when the formerly independent Department of Science which he headed was abolished (see *Nature* 328, 286; 23 July 1987). But with Jones playing a larger role in determining policy, Australian researchers are cautiously optimistic that the importance of science and technology is being recognized by the government.

Tania Ewing

## UNIVERSITIES

# MIT president to quit

## Boston

PAUL E. Gray, president of the Massachusetts Institute of Technology (MIT), announced last week that he will resign in July 1990, the tenth anniversary of his appointment. Following tradition, Gray is expected to become chairman of the MIT Corporation, replacing David S. Saxon, a former president of the University of California, who plans to retire.

Carl M. Mueller, a retired bank executive, will head the search committee for a new president. Mueller was also the head of the committee that hired Paul Gray and he participated in the selection of Gray's predecessor, Jerome Weisner. All five presidents since 1940 have been graduates of MIT.

Gray, an electrical engineer who holds three MIT degrees, has held office during a period when the university has consolidated its stature in areas outside engineering. Biology and biomedical research have particularly thrived in the past decade. Gray helped to negotiate the present relationship between MIT and the well-endowed Whitehead Institute for Biomedical Research.

Seth Shulman



# The hypothetical way of progress

Virginia A. Huszagh and Juan P. Infante

Biologists, and the journals that publish their papers, tend to dismiss theoretical work. Yet it is through ideas, not the mere generation of data, that the course of science is changed.

Is the hypothesis in the biological sciences an endangered species? The history of scientific thought shows that ideas are usually the driving force leading to new discoveries, not the accumulation or cataloguing of data. Many areas of biology, however, including biochemistry, seem stuck with nineteenth-century authoritarian ideology, which favours data production over generating new ideas or questioning stale paradigms. This encourages a kind of 'excellence in mediocrity', which drives off creative minds and attracts researchers who imitate rather than innovate<sup>1</sup>.

The importance of clearly formulated hypotheses which connect data in new ways is demonstrated by the Nobel prize-winning examples of Watson and Crick's hypothesis on DNA structure, Jerne's hypothesis on the origin of immunological diversity, and Mitchell's hypothesis on oxidative phosphorylation. Nevertheless, theoretical work in biology is often viewed as 'speculation'. This is not surprising considering that PhD students in the biological sciences receive little or no exposure to the history and philosophy of scientific thought, making the degree a misnomer. Few biologists can distinguish between speculation, demarcated hypotheses and theories. Even fewer appreciate the need for revolutionary hypotheses; and fewer still can generate them.

## Falsifiable predictions

Understanding the fundamental differences among these concepts is essential if potentially productive proposals are to be distinguished from those that are blind alleys. A hypothesis, to qualify as scientific, must be both grounded (consistent with known experimental observations and fundamental principles) and demarcated (have necessary falsifiable predictions)<sup>2-4</sup>; it thus enables meaningful, directed experimentation. A speculation, on the other hand, is an idea which might appear plausible or even inspiring, but which does not possess the necessary falsifiable predictions needed to make it fertile ground for experimentation. Various so-called 'hypotheses' in biology (often mistakenly labelled as 'theories', which are corroborated hypotheses or related sets of hypotheses) are no more than *post hoc* explanations or non-mechanistic mathematical formulations of data, whose necessary predictions have never been developed or tested. All too

often such "black box"<sup>3</sup> or pseudo-hypotheses are accepted more for socio-political than for scientific reasons<sup>5,6</sup>.

Mitchell's chemiosmotic hypothesis of oxidative phosphorylation<sup>7</sup> is a classic example of a true demarcated hypothesis in biochemistry. It contained only a re-interpretation of evidence already in the literature; it was based largely on theoretical principles<sup>8</sup>; and it proposed several necessary and specific falsifiable predictions. This hypothesis did not arise from a lucky accident, but resulted from an acute appreciation of the inadequacy of the prevailing paradigm, a critical analysis of the literature, and a good deal of reflection and imagination. Mitchell's theoretical work eventually revolutionized the field concerned, although he himself did only a small part of the experimental testing.

What allowed Mitchell to propose his ideas was not that he had access to any new data, but that his mind worked differently. Perhaps it is time that biologists recognized what physicists discovered early this century: that different minds are suited to different aspects of research, and science as a whole benefits by the interplay of theorists and experimentalists. A theorist's strength, and enjoyment, is in proposing testable ideas, whereas that of an experimentalist is in devising tests of such ideas. One cannot thrive without the other; the abilities needed for, and enjoyment derived from, each endeavour are different.

Except in a few isolated areas, those with abilities and a predisposition to do theoretical work find little opportunity or encouragement in biology. Biology is far behind physics in this regard. Indeed, biology is sometimes said to be too 'complex' to allow for general principles; such an attitude is encouraged by observationalists who generate and catalogue vast amounts of data without being able to propose coherent mechanistic explanations of them. This ideology also ignores the power of data that are inconsistent with accepted paradigms, because exceptions are merely seen as part of nature's complexity. The theorist, on the other hand, views inconsistencies as nature's way of saying we are missing something (perhaps the boat).

Only a few biology journals devote themselves largely to non-experimental work (of which hypotheses are only a subset), and they tend not to be widely

read, particularly not by experimentalists. Those who review hypotheses in biology are usually data-generators who do not know the elements of importance and who thus make non-scientific sociologically based objections that are irrelevant to the construction or purpose of a hypothesis.

## Dismissive comments

The theorist frequently encounters such dismissive comments as: "the author presents no new data"; "the author has not done the experiments himself"; or "the author is advised to do the experiments first and return after the hypothesis is 'proven'". Those who take such an attitude are blind to the communal aspect of the scientific enterprise, that the person who proposes an idea need not be the one to test it. Where would physics be if Einstein had had to 'prove' his ideas before publishing them?

Several changes might allow for a more fruitful partnership between theorists and experimentalists in biology. First, more journals could encourage and publish hypotheses. Second, they could provide criteria to define the elements of a legitimate hypothesis. Referees of hypotheses might also be provided with these criteria, to encourage them to make genuine rather than spurious objections. But publication of more testable ideas will achieve little until experimentalists are disposed to seek them out rather than paddling in circles in their own little ponds — it took six years after he published them for Mitchell's ideas to be seriously tested by others. Of course, many so-called 'hypotheses' offer little of the mechanistic insight needed to induce experimentalists to test them; both sides of the equation need improvement. Finally, a few grants to encourage theoretical work and reflection in biology might be in order. After all, only ideas can make sense out of data; data alone do not generate ideas. □

Virginia A. Huszagh and Juan P. Infante are at *The Institute for Theoretical Biochemistry and Molecular Biology, PO Box 4512, Ithaca, New York 14852-4512, USA.*

1. Lane, L.C. *Nature* **335**, 109 (1988).
2. Popper, K.R. *The Logic of Scientific Discovery* 27–135 (Harper & Row, New York, 1968).
3. Bunge, M. *Scientific Research, I. The Search for System* 222–304 (Springer, New York, 1967).
4. Lakatos, I. in *Criticism and the Growth of Knowledge* (eds Lakatos, I. & Musgrave, A.) 91–196 (Cambridge University Press, 1970).
5. Kuhn, T.S. *The Structure of Scientific Revolutions* (2nd edn) 1–210 (University of Chicago Press, 1970).
6. Ninio, J. *Trends biochem. Sci.* **6**, VII-X (1981).
7. Mitchell, P. *Nature* **191**, 144–148 (1961).
8. Mitchell, P. A. *Rev. Biochem.* **46**, 996–1005 (1977).

## Problems of Indian laboratories

SIR—Hair-raising news from an Indian national laboratory has recently horrified the scientific community all over the world. A young bright scientist at such a laboratory (RRL, Bhopal) of the Council of Scientific and Industrial Research (CSIR) was humiliated and harassed, and even physically assaulted to such a degree that he was eventually driven to commit suicide in the laboratory (*Hindustan Times* 15 December 1988).

All commotion subsides with time, and no concrete action will be taken against this state of affairs. Yet in CSIR laboratories in India, bright scientists and especially the young are so humiliated and harassed that they eventually follow one of these three courses: they suffer everything and abandon creativity forever, they escape overseas or they commit suicide.

The fundamental cause of this malady is rooted in the structure of the CSIR laboratories and lies in the autocratic power of the director. A comparison of the working conditions of faculty members in universities and national laboratories in India and in developed countries (say, Britain or the United States) would show that, while Indian research facilities are excellent — CSIR laboratories are luxurious by the yardstick of what India can afford — academic freedom is totally lacking in CSIR laboratories. Instead, the director is all-powerful, and has virtually total control of the future careers of faculty members. There are no checks and balances on the power of the director, as

there are in universities, and no forum in which professional opinions can be freely expressed.

All this leads to the total subjugation of scientists to the director, which breeds sycophancy and corruption. Benefits and prizes go to mediocre but loyal scientists, while the bright ones are tortured.

Those in authority enjoy god-like powers and can ignore all suggestions of protest. Yet the government of India periodically and ritually sets up committees to 'revamp' CSIR, with results that have only cosmetic value. The recent Abid Hussain committee (Mr Hussain himself is not a scientist) remarkably recommended that directors should have even more power, even though they say they do not need it — they have enough already! Occasionally, the dictatorial style of directors is reported in the newspapers, but more often it is suppressed. While the protests of daring honest scientists and international pressure from intellectuals may have some short-term effect, it will be a long time before a research atmosphere can be created for creative scientists in this country. A lot more sacrifice will go into the Ganges before then.

CSIR SCIENTIST

(Name and address withheld)

■ The decision to print this letter, for obvious reasons anonymously, is not taken lightly, but in the knowledge that the circumstances it describes persist in some, but not all, of the laboratories concerned —

EDITOR, *Nature*

## Polio in Israel

SIR—S. C. Arya (*Nature* 336, 708; 1988) refers to the outbreak of paralytic poliomyelitis in Israel last summer, which extended from mid-August to early October and culminated in 15 cases (3.3 per million population) over the seven weeks, the largest number of cases since 1979. Twelve patients lived in the Hadera subdistrict on the Mediterranean coastal plain, and residents of this region were vaccinated in mid-September.

The decision to vaccinate the entire national population (with trivalent oral poliovaccine) was based on the occurrence of three cases outside the Hadera region, widely separated geographically, as well as on serological data suggesting increased susceptibility of the young adult population to type-1 poliovirus. The significance of finding polioviruses in sewage was misunderstood by the Israeli press and others (*Nature* 335, 659; 1988). Health officials searched for wild poliovirus in the sewage as an indicator of where the virus was prevalent, in order to anticipate where the disease might next appear. There was no suggestion of transmission of disease through sewage.

The Health Ministry's decision to vaccinate only those under age 40 was based on the epidemiology of poliomyelitis in Israel: between 1971 and 1987, only one of 170 cases of poliomyelitis was a person 40 years old. Arya is right in saying that most people aged 40 or over had never been vaccinated, but it was felt that almost everyone who had lived through the epidemics of the 1950s had been immunized by natural exposure to the virus.

The ministry's decision was vindicated after the fact: the outbreak was stopped in its tracks by the vaccination campaign. Furthermore, analysis of blood samples taken from adults from two subdistricts before vaccination showed that 145 of 146 people aged 30 years and older (99.3 per cent) had neutralizing antibody titres of at least 1:4 against all three poliovirus types.

PAUL E. SLATER

Department of Epidemiology,  
Ministry of Health,  
Jerusalem, Israel

RACHEL HANDSHER

Central Virology Laboratory,  
Sheba Medical Center,  
Tel Hashomer, Israel

## Kidney diplomacy

SIR—Your conclusion (*Nature* 337, 393; 1989) that the shortfall in donor kidneys might be made good by better international cooperation calls for serious consideration. In this sphere as in many others, two European maps have been drawn, with two separate international organizations, Eurotransplant for Western Europe and Intertransplant for the COMECON countries. This division has, to our knowledge, been breached only once, when in 1984 a Moscow surgeon sent a kidney for which no recipient had been found to Britain, where it was transplanted into a 59-year-old man (*The Times* 28 May 1985). Surely this initiative could in the present climate of increasing cooperation be taken up and extended.

STEWART BRITTEN  
(Hon. Secretary)

UK-USSR Medical  
Exchange Programme,  
9d Stanhope Road,  
London N6 5NE, UK

## RU 486 abortions

SIR—Having read Peter Coles's article (*Nature* 335, 486; 1988), I think that I must have expressed myself badly when talking to him.

Strictly speaking, there has been no case of a malformed baby being born to a mother who has taken RU 486 during pregnancy. The case of fetal malformation reported to the French health authorities was discovered after a termination of pregnancy at 18 weeks from conception for a very severe oligoamnios, in a woman who had taken RU 486 and whose pregnancy continued. The embryological examination did not allow any conclusions to be drawn regarding the role of RU 486.

Given the fact that, in the case of a failure of RU 486, even when associated with a prostaglandin, the pregnancy may indeed continue to develop and that the fetus is therefore exposed to a risk of malformation — as suggested by studies in the rabbit but not confirmed in the mouse and the rat — it is imperative to suggest a termination of pregnancy by another method.

It is stipulated that the woman shall be informed of these risks and should sign, before administration of RU 486, a letter of consent in which she declares that she has understood the limitations of the method and in particular the potential risk of embryotoxicity should the pregnancy be continued despite her having taken RU 486 followed by prostaglandin.

R. HENRIEN

Assistance Publique,  
Hôpitaux de Paris,  
Clinique Universitaire Port-Royal,  
123 boulevard Port-Royal,  
75014 Paris, France



# The biggest greenhouse still intact

The CFCs are plainly headed for extinction, but those at this week's London conference should not on that account believe the greenhouse effect has been banished.

QUITE what last weekend's conference on the ozone inhibitors accomplished is a matter for conjecture, but there is a high chance that many of those present believed they were assisting decisively in the preservation of the familiar pattern of the Earth's climate. In some ways, of course, they were. The chlorofluorohydrocarbons, sometimes called chlorofluorocarbons or CFCs, are also greenhouse gases, with absorption spectra scattered throughout the infrared.

More intricate processes must also be considered. To the extent that increasing concentrations of CFCs may diminish the average concentration of stratospheric ozone, and because stratospheric ozone is an element in the determination of the Earth's radiation balance, there is also a possibility that the increasing concentration of the CFCs may undermine, in more subtle ways, the climate as we know it. But, also, it may not.

What we all need, in circumstances such as these, is some kind of yardstick. Here is one — the average external radiation flux reaching the Earth, roughly  $236 \text{ W m}^{-2}$ . At equilibrium, that is the power that enters the atmosphere with a temperature equivalent to that at the surface of the Sun; it is also the power that escapes from the top of the Earth's atmosphere. The yardstick, of course, is huge. Each 2-m patch on the Earth's surface must receive and then get rid of enough power to keep a person warm indefinitely.

The notion that  $\text{CO}_2$  may influence this balance is antique; John Tyndall of Manchester seems to have been the first to draw attention to the possibility that the ubiquitous product of combustion may affect the radiation balance (*Phil. Mag.* **22**, 161; 1861). In reality, we now know that, were it not for  $\text{CO}_2$ , life on the surface of the Earth would not be possible. During the last glaciation, the concentration of  $\text{CO}_2$  was low, as was the average temperature.

The surface of the Earth, with an average temperature of 288 K, emits radiation (mostly in the infrared) at a power of  $390 \text{ W m}^{-2}$ , more than half as great again as the incoming flux of solar radiation. There follows an exchange of energy within the troposphere, which we call weather, that evens out the temperature. If it were otherwise, the Earth would be like Mars, unable to sustain life at any

season.

These are also the reasons why we have a troposphere, a layer of the atmosphere within which the temperature decreases with increasing altitude. The Earth's atmosphere is literally a blanket within which the passage of infrared radiation is everywhere impeded by absorption and re-radiation. The temperature gradient is a measure of how much of the energy is diverted to meteorological processes.

Lovelock's Gaia hypothesis notwithstanding, there is no simple reason why this lucky state of affairs should persist indefinitely. There are two ways of regarding the recognition that the CFCs are greenhouse gases (infrared absorbers) as well as destroyers of ozone. One is that their accumulation may increase the temperature uncomfortably, another is that they would have seemed a godsend to those who, just twenty years ago, believed that the then downward trend of temperature presaged another glaciation. Why not make a few million tonnes of the stuff and pump it into the atmosphere, they might have said?

This is where the yardsticks come in. First,  $\text{CO}_2$  absorbs infrared radiation principally on the long wavelength side of  $12 \mu\text{m}$ , while the CFCs in their role as infrared absorbers are active in regions scattered on the high-frequency side of that. In other words, they occupy what would otherwise be a window of wavelengths transparent to the relatively high-frequency infrared.

Second, molecule for molecule, CFC molecules are much more efficient absorbers than  $\text{CO}_2$ . So much should be self-evident: molecules of both kinds absorb in the infrared because they vibrate, but CFCs are more significantly polar molecules in which dipolar interactions with radiation must be more significant. Some CFC molecules are said to be 10,000 times as efficient as  $\text{CO}_2$  molecules at absorbing infrared radiation of an appropriate frequency. So should not their accumulation in the atmosphere bring trouble?

The question is not whether, but how much. And the same is true for  $\text{CO}_2$ . It would be an assault on rational expectations to suppose that increasing the concentration of a greenhouse gas would make the greenhouse less effective, but Le Chatelier's principle would similarly lead

one to expect abatements of simple expectations which are comparable in size but necessarily smaller.

Broadly speaking, most calculations carried out so far are in agreement with one other. A doubling of the concentration of atmospheric  $\text{CO}_2$  (to a total concentration of 600 p.p.m.) would increase the surface temperature of the Earth by about  $3^\circ \text{C}$  Centigrade, or by roughly six times as much as a doubling of the concentrations of the two principal infrared absorbers among the CFCs (see Ramanaathan *et al. Rev. Geophys.* **25**, 1441; 1987). These estimates do not allow for the obvious feedbacks in the system, among which the more plentiful formation of clouds with increasing temperature must be conspicuous. (Ordinarily, clouds will cool, although there may be circumstances in which they do the opposite.)

The reason for worrying about CFCs as greenhouse gases is that their chemistry is different, perhaps significantly so. The production of  $\text{CO}_2$  by combustion is now so great that its concentration would double every half-century, but the observed increase of concentration is only half as quick, presumably because of interaction with the oceans and the biosphere (which should on paper be luxuriating).

By contrast, some CFCs are believed to have half-lives measured in centuries, but on the assumption that they are removed from the atmosphere only by the photolytic processes that lead to the destruction of ozone, while their rate of release to the atmosphere has recently been increasing much more quickly than that of  $\text{CO}_2$ . In principle, CFCs could indeed be a bigger worry than  $\text{CO}_2$  a few decades from now, which is why it is creditable that their removal has been given such attention.

Even so, there are some persisting doubts, of which the chief must be that CFCs may not be as chemically stable as reputed. What of the possibility that the great Antarctic ozone hole is a place where some of this material is washed out onto the ice-cap? But there is also some certainty that, whatever happens to the CFCs, the concentration of  $\text{CO}_2$  will continue inexorably to increase, whatever is done to amend the Montreal Protocol at Helsinki in May. That, in the circumstances, is the greenhouse that matters.

John Maddox



# Genomic imprinting and embryonal tumours

Wolf Reik and M. Azim Surani

SOME alleles in the mammalian genome can function differently depending on whether they come from the mother or the father. Functional differences between homologous alleles are caused by genomic imprinting, an epigenetic marking process which has important consequences for embryonic development<sup>1</sup>. The presence of two maternal or two paternal copies of particular chromosomes, for instance, can result in abnormal growth and development in mice<sup>2</sup>. The unexpected finding of preferential loss of maternal chromosomes in embryonal tumours described by Toguchida *et al.* on page 156 of this issue<sup>3</sup>, and by Schroeder *et al.*<sup>4</sup> and Mannens *et al.*<sup>5</sup> in two earlier reports, now suggests a link between imprinting and malignant growth in man. These observations raise the exciting possibility that the different function of maternal and paternal alleles caused by genomic imprinting is a crucial event in the genesis of recessive tumour syndromes.

An increasing number of cancer syndromes are being discovered to result from recessive mechanisms in somatic cells, as discussed recently in News and Views<sup>6</sup>. Two mutant alleles of a putative tumour-suppressor gene are produced by first mutating one allele, followed by the loss of the second functional allele, which is often caused by mitotic recombination events. Tumour tissues become homozygous for markers located on the chromosome that harbours the first mutation. In heritable cases, the chromosome retained in the tumour is of course derived from the affected parent. In sporadic cases, however, most of which are thought to be a consequence of somatic mutations,

the maternal and the paternal gene would be expected to have an equal chance of carrying the mutation. The new findings of Schroeder *et al.*<sup>4</sup>, of Mannens *et al.*<sup>5</sup> and of Toguchida *et al.*<sup>3</sup> reported in this issue show that this expectation was wrong. In most sporadic cases of both Wilms tumour<sup>4,5</sup> and osteosarcoma<sup>3</sup> it is the paternal chromosome that is retained in the tumour, the maternal one being lost.

Wilms tumour and retinoblastoma/osteosarcoma are among the first tumours for which a recessive mechanism has been suggested<sup>7-9</sup>. Wilms tumour (or nephroblastoma) is assumed to result from a homozygous mutation of a gene located on chromosome 11p13 and perhaps on 11p15 (ref. 5). The loss of heterozygosity for extended regions of 11p and indeed of the entire chromosome 11 are frequently observed. Interestingly, several other tumours such as rhabdomyosarcoma and hepatoblastoma also show loss of heterozygosity at 11p, suggesting that the same gene can result in different neoplasms depending on the embryonic lineage in which its loss occurs.

Similarly, retinoblastoma arises from a homozygous mutation of a gene on chromosome 13q14. Patients that survive the heritable form of retinoblastoma show a greatly increased incidence of osteosarcoma. Loss of heterozygosity for 13q is also observed in sporadic cases of osteosarcoma, and the recently isolated retinoblastoma gene (RB) has been found to be deleted in some osteosarcomas. As in the case of Wilms tumour, it seems that homozygous loss of the RB gene can result in retinoblastoma if the loss of the second allele occurs in the embryonic retina, and

in osteosarcoma if the second mutation happens to occur in bone tissue.

One caveat, although unlikely, does apply: it is possible that the same gene, but different mutations, are responsible for retinoblastoma and osteosarcoma, respectively. Homozygous deletions of the RB gene have been observed in both tumours, however, suggesting that if the same gene is responsible, the common denominator should be loss of function.

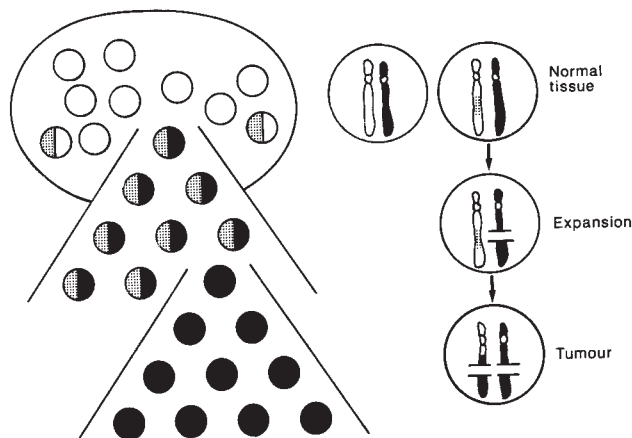
In sporadic cases of Wilms tumour and retinoblastoma/osteosarcoma, the first mutation can occur either in maternal or paternal gametes or in somatic cells of the embryo. The second mutation must be somatic, and most frequently this involves mitotic nondisjunction or recombination and leads to the observed loss of heterozygosity. It is not difficult to envisage that mutation rates differ in spermatogenesis and oogenesis; indeed, a preponderance of *de novo* structural abnormalities of paternal origin has been noted<sup>10</sup>. If, however, germinal mutation is excluded and there is still a bias in which parental chromosome is retained in the tumour, a different explanation is needed.

Toguchida *et al.*<sup>3</sup> take an interesting approach to this problem by deciding to investigate parental origin of chromosome 13 in sporadic osteosarcoma. They argue that because none of their patients has retinoblastoma, the first mutation should be somatic and should be segregated into bone tissue but not into retina. Of 13 sporadic osteosarcomas in which they find loss of heterozygosity for 13q markers, 12 retain the paternal chromosome. The authors exclude three cases in which the first mutation could have been on the maternal chromosome, and so the final ratio of paternal to maternal retention in their patients is 9:1.

Schroeder *et al.*<sup>4</sup> report five cases of sporadic Wilms tumour; in all five they demonstrate loss of maternal alleles in the 11p region. Most bilateral tumours and tumours associated with aniridia and genitourinary anomalies are thought to be caused by germinal mutation, so these have to be excluded. It is only unilateral tumours that are mainly caused by two somatic events. Schroeder *et al.* therefore exclude two of their patients, but include two unilateral cases from the published literature where loss of maternal alleles was reported. Hence, five out of five tumours in which somatic genesis was likely retain the paternal 11p region. Mannens *et al.*<sup>5</sup> find that of five tumours homozygous for 11p markers, four retain the paternal alleles. Unfortunately, they do not say how many of these were unilateral tumours.

Schroeder *et al.* offer no explanation for their findings; Mannens *et al.* invoke the possible involvement of genomic imprinting; and Toguchida *et al.* explicitly state that the susceptibility to mutation of maternal and paternal alleles could be

Loss of maternal alleles in tumours. Most cells in the target tissue have maternal (open) and paternal (filled) chromosomes, both of which express either the RB gene or the Wilms gene. On some maternal chromosomes, however, these genes are repressed by methylation imprinting (stippled area). If the first mutation happens to be on a paternal chromosome opposite the relatively inactive maternal allele, this population of cells will expand. This expansion will neither occur with the first hit on a paternal chromosome opposite an active maternal allele nor with the first hit on a maternal chromosome. The expanded population now represents a much increased target for the second event — loss of the maternal allele (by somatic recombination in this case). Note that if the second mutation does not occur, the initial focus of growth may not progress into a tumour.



different because of differences in epigenetic modification resulting from genomic imprinting. Certainly, we must look for a mechanism that either produces differences in mutation between parental alleles, or that selects for differences between them. The best molecular evidence for an imprinting process comes from experiments with transgenic mice, where several groups have reported examples of differences of DNA methylation between parental alleles<sup>11</sup>. How there could be selective somatic mutation on the paternal allele of the RB gene is not clear from these studies. If the paternal allele is more methylated in most somatic cells, the increased frequency of mutation by deamination of 5-methylcytosine could contribute to selective mutation of the paternal gene. Both the methylation status of the paternal RB allele and the frequency of mutation at CpG dinucleotides can now be analysed with the RB gene in hand.

On the basis of the observations in Wilms tumour, Wilkins<sup>12</sup> proposed that a transforming gene (TR) linked to the recessive Wilms gene on chromosome 11 is imprinted. Inactivation of the Wilms gene product derepresses this transforming gene and results in development of the tumour. However, the maternal allele of TR is rendered inactive by methylation imprinting and so it is only the combination of the inactive Wilms gene and the active (paternal) TR on the paternal chromosome that results in a tumour. Two assumptions are implicit in this model, and they should become testable. First, inactivation by imprinting of the maternal TR allele should be observed in most, if not all, cells of the embryonic kidney. Second, whenever the first mutation is in the maternal Wilms gene, a tumour should arise when somatic recombination between the Wilms locus and TR produces a hybrid chromosome with the Wilms mutation and the active (paternal) transforming gene. If this is correct, the information on location of the recombination breakpoints could also be used to map the position of the proposed TR gene.

Finally, we wish to draw attention to an alternative explanation based on a suggestion originally made by Sapienza *et al.*<sup>13</sup>. Based on various experimental observations, these authors proposed that imprinting by methylation is not retained in most cells in an embryo. In this explanation, tissues in the embryo are composed of many cells in which both alleles are active, and of rather a few cells that retain methylation differences between specific parental alleles. If we extrapolate from the findings with transgenes in the mouse, then in these cells the maternal allele is likely to be relatively inactive because of hypermethylation (see figure). If the first mutation in the Wilms gene or the RB gene occurs on the paternal copy in these cells, and is segregated into the correct target

tissue, a focus of tumour cells will appear in kidney or bone. This focus greatly expands the number of target mitoses for the second event, through which the remaining allele is lost. The remaining (maternal) allele is not necessarily completely repressed by methylation; indeed varying degrees of reduction in activity can be envisaged that may even change during the growth of the focus. So, what remains in this model is the possibility that the 'second hit' does not happen. Perhaps this could explain a frequent observation in tumour biology, one also true for retinoblastoma and Wilms tumour — the finding of tumour-like lesions or foci in these tissues that have regressed (usually a chance discovery on autopsy).

Much of the present uncertainty about the true mechanism underlying the unequal representation of parental chromosomes in these tumours should soon be removed by the techniques of molecular biology. At least now that the RB gene has been cloned, various questions raised by the observations we have

described here can now be addressed by looking closely at mutations, at transcription, and at methylation. The most exciting prospect is that the same principle may operate in many, if not all, recessive tumour syndromes. Further analysis of this principle may reveal more about the significance of genomic imprinting in the control of normal and malignant cell differentiation. □

Wolf Reik and M. Azim Surani are in the Department of Molecular Embryology, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK.

1. Solter, D. A. *Rev. Genet.* **22**, 127–146 (1988).
2. Cattanaach, B. M. & Kirk, M. *Nature* **315**, 496 (1985).
3. Toguchida, J. *et al.* *Nature* **338**, 156–158 (1989).
4. Schroeder, W. T. *et al.* *Am. J. hum. Genet.* **40**, 413 (1987).
5. Mannens, M. *et al.* *Hum. Genet.* **81**, 41–48 (1988).
6. Ponder, B. *Nature* **335**, 400–402 (1988).
7. Knudson, A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 820 (1971).
8. Knudson, A. G. & Strong, L. C. *J. natn. Cancer Inst.* **48**, 313–324 (1972).
9. Comings, D. E. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3324–3328 (1973).
10. Magenis, R. E. *Am. J. hum. Genet.* **42**, 529–533 (1988).
11. Surani, M. A. H. *et al.* *Trends Genet.* **4**, 59–62 (1988).
12. Wilkins, R. F. *Lancet* **i**, 329–331 (1988).
13. Sapienza, C. *et al.* *Prog. Nucleic Acids Res. molec. Biol.* (in the press).

## An anniversary for atmospheric ozone

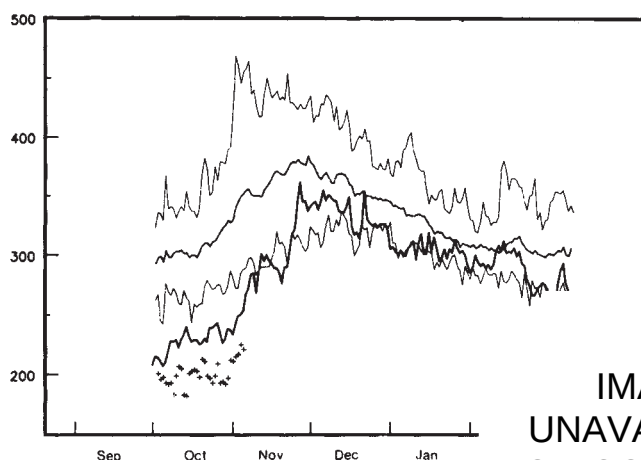


IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

G. M. B. DOBSON, who pioneered atmospheric ozone measurements, was born on 25 February 1889. The photograph was taken while Dobson was a member of the Meteorological Office Research Committee in 1955. His interest in ozone was aroused by his studies of meteor trails in the atmosphere, which led to his discovery of a high-temperature region at a height of about 50 kilometres. His realization that the source of energy for this warm layer is the absorption of ultraviolet radiation by ozone inspired him to spend the rest of his life studying atmospheric ozone. Dobson developed a technique for measuring ozone in which a photoelectric spectrophotometer is used to compare the absorption in the solar spectrum resulting from ozone at two ultraviolet wavelengths in the absorption band of ozone by allowing the radiation of each wavelength to strike the photocell. In the 1930s Dobson set up a network of ozone-measuring stations throughout the world, many of which are still in existence. Indeed, the Antarctic ozone hole was discovered by a team from the British Antarctic Survey working at Halley Bay in Antarctica (*Nature* **315**, 207; 1985; see graph) using a Dobson instrument to measure ozone. Satellite measurements of ozone began only in the late 1970s and so long-term changes and trends in atmospheric ozone concentration can be assessed only using the older, ground-based data. The satellite data, while providing more comprehensive global coverage, are prone to suffer long-term drifts in instrument calibration when in orbit. To compensate for this effect, therefore, the satellite data are normalized against the Dobson data.

Philippa Lloyd



# Stress-induced segregation

Charles G. Sammis

EARTHQUAKES and topographic uplift of the Earth's surface are generally recognized to be the direct results of slow convective flow in its hot interior. The solid rock of the Earth's mantle flows by a combination of dislocation and diffusion mechanisms. Rocks from the upper mantle are occasionally brought to the surface either by deeply rooted volcanoes or by obduction (overthrusting) associated with the collision of rigid surface plates. Microscopic studies of such rocks indicate that mantle flow is dominated by the generation and motion of dislocations: the density of dislocations and the size of associated structures such as subgrains are routinely used to estimate the magnitude of stress associated with mantle flow<sup>1</sup>. But on page 141 of this issue<sup>2</sup>, K. Ozawa presents the first evidence that refractory oxide minerals in upper-mantle rocks may deform in a creep regime where strain is produced by the direct diffusion of atoms from grain boundaries under high stress to those under lower stress.

Ozawa's evidence is the observed chemical segregation of aluminium and

chromium in elongate grains of chromian spinel —  $(\text{Cr}, \text{Al}, \text{Fe}^{3+})_2 (\text{Mg}, \text{Fe}^{2+})\text{O}_4$  — in ultramafic upper-mantle rock. Although such a chemical unmixing of initially homogeneous solid solution oxides has been predicted to occur during diffusion creep<sup>3</sup>, it has never been observed either in the laboratory or in nature. Ozawa's observations suggest that stress-induced 'kinetic demixing' produces easily measured compositional gradients under upper-mantle conditions and may therefore form the basis of a new technique to measure the orientation and magnitude of flow stresses in the Earth's mantle.

The most surprising result of Ozawa's study is the magnitude of the observed zonation which, for  $\text{Al}_2\text{O}_3$ , is about 10 per cent across a grain: laboratory experiments have failed to detect any measurable demixing<sup>4,5</sup>. As discussed below, this may be a consequence of nature's ability to access more favourable experimental conditions at strain rates too low to simulate on a laboratory timescale. The stress-induced demixing of solid-solution oxides also has potential consequences in

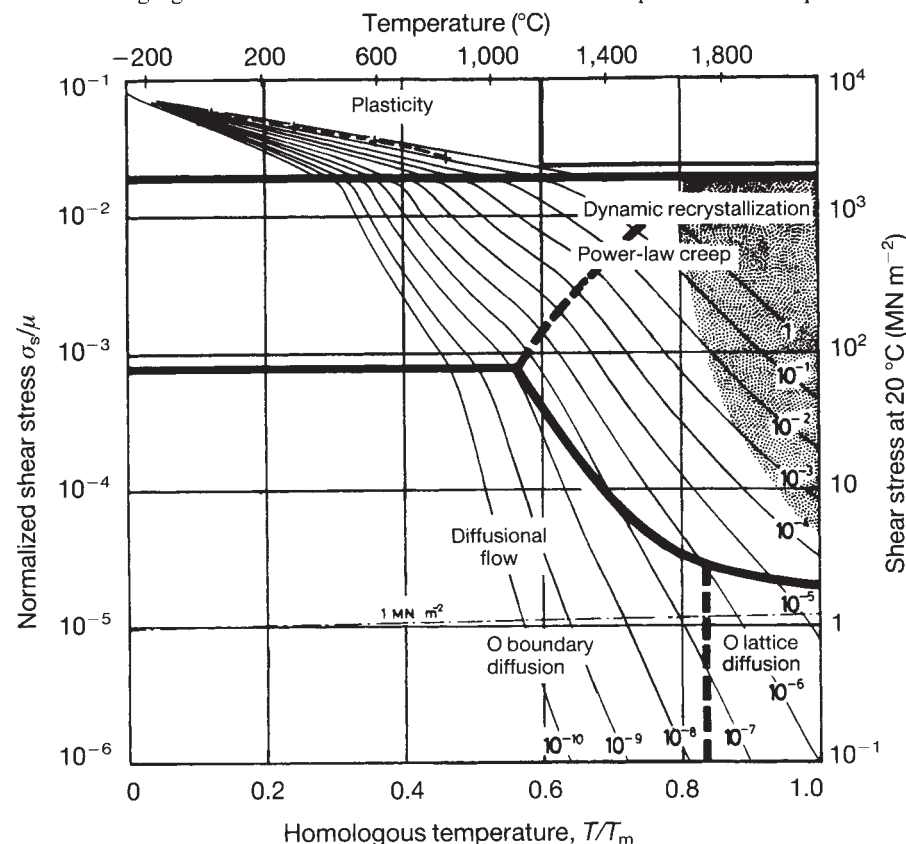


FIG. 2 A stress-temperature deformation mechanism map for magnesium aluminate spinel with a grain size of  $10\ \mu\text{m}$ . Heavy lines form the boundaries where adjacent mechanisms make an equal contribution to the strain rate. Light lines are contours of constant strain rate (from  $1\ \text{s}^{-1}$  to  $10^{-10}\ \text{s}^{-1}$ ). Flow is accommodated in the 'power law creep' field by the motion of dislocations and in the 'diffusional flow' field by the diffusion of individual atoms. Note the reduction in the upper shear-stress boundary of the diffusional flow regime as  $T$  approaches the melting temperature,  $T_m$ . (From ref. 7.)

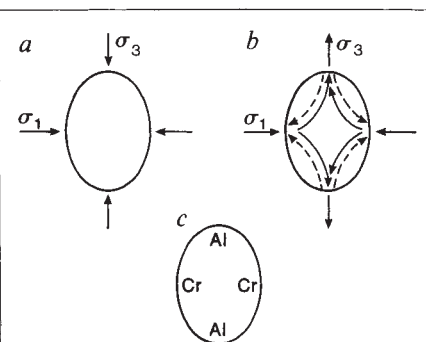


FIG. 1 Diffusion creep occurs when unequal stresses are applied to a grain ( $a$ ;  $\sigma_1$  and  $\sigma_3$  are the largest and smallest principal stress, respectively). Diffusion arises from the deviatoric stresses ( $b$ ), defined as the differences  $\sigma_{1,3} = \sigma_{1,3} - \sigma_m$ , from the mean stress  $\sigma_m$ . Vacancies flow from regions of tension to regions of compression; atoms in the contrary sense. If diffusion is not limited by oxygen flow, and Al diffuses faster than Cr, the grain boundaries become enriched relative to the bulk as shown in  $c$ .

ceramics and materials science to the extent that such zonation degrades the physical properties of fine-grained ceramics in long-term, high-temperature, low-stress applications (see the recent News and Views article<sup>6</sup> by R.W. Cahn).

Stress-induced kinetic demixing of an oxide solid solution  $(\text{A}, \text{B})\text{O}$  can occur under the following conditions<sup>3</sup>: first, the stress level must be low enough that deformation occurs by atomic diffusion; second, cation diffusion must be rate-limiting (that is, O diffusion must be more rapid than either A or B); and third, the diffusivity ( $D$ ) of one of the cations must be significantly greater than the other. If  $D_A > D_B$ , the concentration of AO will be enriched at grain boundaries in tension and BO will be enriched at grain boundaries in compression (Fig. 1).

Ozawa argues that these conditions were met in his naturally deformed spinel grains. He suggests that oxygen diffuses along grain boundaries at a rate comparable to cation lattice diffusion, and thus is not rate-limiting. Although the mobilities of Al and Cr in chromian spinel are not known, Ozawa argues that Al probably diffuses more quickly: it is faster in  $\text{MgO}$ . Also the melting temperature of Al-spinel is lower than that of Cr-spinel, and diffusivity is inversely correlated with melting temperature<sup>1</sup>. Faster diffusion of Al is consistent with the observed enrichment of  $\text{Al}_2\text{O}_3$  at the elongated (more tensile) grain boundaries.

The steady-state compositional difference between tensile and compressive boundaries is given by<sup>3</sup>

$$\Delta X^o = \frac{X^o(1-X^o)(r-1)}{(1-X^o)r+X^o} \times \frac{\Delta\sigma V_m}{RT}$$

where  $X^o$  is the mole fraction of  $\text{Al}_2\text{O}_3$ ,  $r$  is the ratio of diffusivities  $D_A/D_B$ ,  $\Delta\sigma$  is the



deviatoric stress and  $V_m$  is the molar volume ( $R$  and  $T$  are the gas constant and temperature, respectively). Laboratory attempts to observe stress-induced demixing in oxides have not been successful. For yttria-stabilized zirconia, failure to observe demixing was attributed to similarity of the diffusion rates of the two cations ( $r=1$ ) and to a possible rotation of the grains during deformation<sup>4</sup>. Attempts to observe demixing in the solid solution (Co,Mg)O were frustrated by the observation that the diffusion creep was rate limited by oxygen diffusion<sup>5</sup>.

Ozawa did not attempt to evaluate the equation for his chromian spinel because the difficulties of Al and Cr in the spinel lattice are not available. However, it is still interesting to evaluate it in the limit  $D_{Al} \gg D_{Cr}$  in which case  $r \gg 1$ . In this limit

$$\Delta X^s \approx X^o \times \frac{\Delta \sigma V_m}{RT}$$

Using  $X^o=0.5$ ,  $V_m=10^{-4} \text{ m}^3 \text{ mol}^{-1}$ , and  $T=1,200 \text{ K}$  as a typical upper-mantle temperature, Ozawa's observed compositional gradient of  $\Delta X^s=0.1$  corresponds to a deviatoric stress of  $\Delta \sigma=10$  megapascal (MPa). A higher stress would be required if the diffusivities were more nearly equal.

To help put this stress in some context, consider the deformation map for  $\text{MgAl}_2\text{O}_4$  spinel (Fig. 2)<sup>7</sup>. Although this map is for a smaller grain size of  $10 \mu\text{m}$ , the following arguments still apply to Ozawa's larger grains. First, note that the point 10 MPa and 1,200 K lies in the diffusion creep field when oxygen diffusion occurs along grain boundaries (as postulated by Ozawa). More importantly, note that the strain rate under these conditions is between  $10^{-8}$  and  $10^{-9} \text{ s}^{-1}$ . This is more than an order of magnitude slower than the lowest strain rates obtained in laboratory attempts to observe demixing<sup>5</sup>.

To achieve measurable strain-rates, laboratory experiments are generally run nearer the melting temperature (in the lower right-hand corner of Fig. 2). Note, however, that near the melting temperature, diffusion creep rate is limited by the lattice diffusion of oxygen, which is one of the conditions which precludes demixing. Even if one were careful to stay in the grain-boundary diffusion field at laboratory strain rates near  $10^{-7} \text{ s}^{-1}$ , the diffusion field shrinks to lower stresses as one approaches the melting point. The combination of larger  $T$  and smaller  $\Delta \sigma$  in

the simplified equation would lead to a smaller compositional gradient in such a laboratory experiment.

It thus seems that the reason that stress-induced kinetic demixing is easily observable in naturally deformed spinels, but has been so elusive under laboratory conditions, is that nature has time to run the experiment at very low strain-rates in the grain boundary diffusion regime. The corresponding higher stress and lower

## EVOLUTIONARY ECOLOGY

# Choosing parasite-free mates

Andrew Pomiankowski

Do sage grouse males congregate at display grounds (leks) and strut before prospective mates to advertise their freedom from parasites? Hamilton and Zuk a few years ago proposed that resistance to parasites may be reflected in secondary sexual characters, allowing females to pick mates resistant to common and harmful parasites and so pass on resistance genes to their offspring<sup>1</sup>. Within-species tests of this hypothesis that justify several of its premises — male ornaments reveal parasite loads, females prefer less parasitized males as mates, resistance is heritable, and only harmful parasites are involved in mate choice — were reported at a recent meeting\*.

In red jungle fowl (*Gallus gallus*), ornamental characters are particularly good indicators of parasite infection (M. Zuk, University of New Mexico, Albuquerque). Given a choice of males, females preferred healthy, uninfected controls to males that have been fed with intestinal roundworm egg inocula as chicks. Control birds had redder hackle feathers, larger combs and greater iris brightness and colour saturation.

By contrast, non-ornamental traits, such as body weight, bill size and saddle feathers, showed no significant difference between parasitized and control birds. This finding does not support the idea that females mate with less-parasitized males merely as an incidental side-effect of their mate preferences. If this were the case, then some non-ornamental traits should also be found to vary closely with parasite burden. Zuk plans to investigate whether ornaments in jungle fowl are insensitive to other non-biotic stresses and strains. If so, this would demonstrate that ornaments are specific indicators of parasite resistance and not just of general viability. An explanation is also needed for why colour is an honest indicator of parasite loads.

There are two ways in which ornaments might indicate differences in parasite loads<sup>2-5</sup>. First, the condition of the orna-

temperatures produce readily observable compositional gradients. Such chemical demixing in grains of refractory-oxide solid solutions seem to have the potential to provide an important compliment to current stress-determination techniques which use dislocations in the lower melting components of upper-mantle rocks. □

Charles G. Sammis is in the Department of Geological Sciences, University of Southern California, Los Angeles, California 90089, USA.

ment itself may reveal parasitism. For example, louse-infested male sage grouse (*Centrocercus urophasianus*) with more lice secure fewer copulations in arena experiments (M. Boyce, University of Wyoming, Laramie). But females cannot directly assess lice loads and show no greater preference for deloused males. However, they appear to reject lousy males because their airsacs, inflated during courtship, have many small black

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

*P. colchicus* — testing for natural resistance. haematomas caused by lice infestation. Males with artificial haematomas applied with a pen are also rejected by females.

Second, females can judge parasite loads if ornament display is condition-dependent — for instance, if only healthy males can court successfully. Male sage grouse with malaria copulate less frequently, and generally with hens in poor condition (Boyce). Every few days they suffered from malarial relapse and were probably too weak to attend the lek. This may explain why females visit the lek for 2–3 days in succession and require males to do the same before mating with them.

A very different picture emerges from sage grouse leks in Mono County, California, studied by R. Gibson (University of California, Los Angeles). Gibson found no malaria in these populations, only infections of *haemoproteus* blood parasites, which showed no correlation with any measure of male mating success.

1. Poirier, J.-P. *Creep of Crystals* (Cambridge University Press, 1985).
2. Ozawa, K. *Nature* **338**, 141–144 (1989).
3. Dimos, D., Wolfenstine, J. & Kohlstedt, D.L. *Acta metall.* **36**, 1543–1552 (1988).
4. Dimos, D. & Kohlstedt, D.L. *J. Am. Ceram. Soc.* **70**, 531–536 (1987).
5. Dimos, D., Kohlstedt, D.L. & Schmalzried, H., *J. Am. Ceram. Soc.* **71**, 732–741 (1988).
6. Cahn, R.W. *Nature* **336**, 19 (1988).
7. Frost, H.J. & Ashby, M.F. *Deformation-Mechanism Maps*. (Pergamon, Oxford 1982).

\* *Parasites and Sexual Selection*, American Society of Zoologists, San Francisco, 27–30 December 1988. Organized by G. Hausfater and R. Thornhill.

However, this may result from the low prevalence of *haemoproteus*: less than 6 per cent of cocks had the parasite. The small population size and isolation of the Mono County leks probably explains the absence of serious parasites there.

Several other studies apparently show that parasite-free males are not the preferred mating partners. A particularly clear example was a three-year study of grey tree frogs (*Hyla versicolor*). No consistent negative correlation was observed between call duration, the primary cue in female choice, and helminth or nematode infestation (G. Hausfater, University of Missouri, Columbia). But this does not refute the parasite hypothesis because none of these parasites had more than minor physiological effects on tree frogs. There is little or no benefit for a female picking a mate free of parasites that are, in general, non-debilitating and non-lethal.

One area of the parasite hypothesis that remains untested is the claim that heritable variation in parasite resistance is

maintained by coevolutionary cycles between host and parasite genotypes. But heritable resistance has been reported in ring-necked pheasants (*Phasianus colchicus*; see figure). Chicks were reared either under hygienic conditions, being fed low levels of an anti-protozoan drug (coccidia being the main parasite) or under unhygienic, untreated conditions to allow full exposure to parasites (N. Hillgarth, University of Oxford). Mortality in the untreated group was twice as high as in the treated group, but offspring of untreated parents survived better than those of treated parents. Having been subject to the rigours of parasite selection, the untreated parents ought to have higher natural resistance, which they did. □

Andrew Pomiankowski is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Hamilton, W. D. & Zuk, M. *Science* **218**, 384–388 (1982).
2. Pomiankowski, A. N. *Oxford Surv. Evol. Biol.* **7**, 136 (1988).
3. Read, A. *Trends Ecol. Evol.* **3**, 97–102 (1988).
4. Maynard Smith, J. *J. theor. Biol.* **115**, 1–8 (1985).
5. Andersson, M. *Evolution* **40**, 804–816 (1986).

## THERMOMETRY

# Ironing out the standard scale

R. L. Rusby

FROM 1 January 1990 the boiling point of water at standard pressure will no longer be taken to be 100 °C exactly. For many people this may be the most striking consequence of the decision, taken recently by the International Committee for Weights and Measures, to revise the International Temperature Scale. The new scale will incorporate many advances in both fundamental and practical thermometry which have been made since the last revision in 1968.

Since 1954, the kelvin has been defined by assigning the value 273.16 K to the temperature of the triple point of water — the unique point at which ice, water and water vapour co-exist in equilibrium. Celsius temperatures have been defined by subtracting 273.15 from the kelvin equivalents. It follows from these definitions that the values of all other temperatures, the boiling point of water included,

must be determined using thermodynamic instruments such as the constant-volume gas thermometer. This is a laborious process: as Callender noted in 1899: "it is impossible for those who have never worked with a gas thermometer to realise the extent of its shortcomings", and no practitioner today would disagree.

On the other hand there are many devices which respond to temperature changes with good precision and reproducibility, and these can be used as practical thermometers even if the values obtained are not accurate in the thermodynamic sense. The objective of the International Temperature Scale, of which Callendar was a pioneer, has always been to combine this precision and reproducibility with as good thermodynamic accuracy as can be achieved, and so provide a scale to satisfy all but the most stringent requirements. The first Inter-

national Temperature Scale was adopted in 1927, and revisions were undertaken in 1948 and 1968 to improve its accuracy, to introduce more reproducible fixed points, and to extend the temperature range. The new scale, to be known as ITS-90, is a continuation of this process.

For the first time primary techniques other than gas thermometry have contributed to the

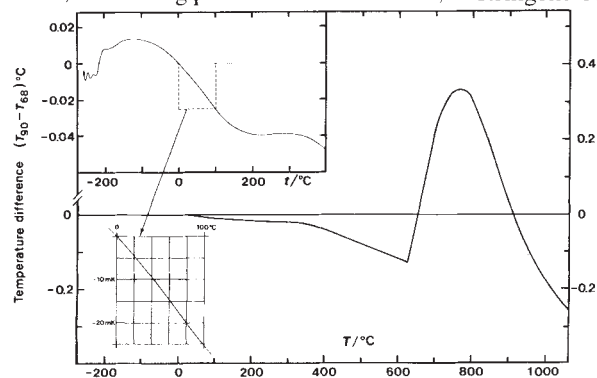
realignment of temperature values. Thus values above 500 °C will be based principally on spectral radiation pyrometry using Planck's radiation law. Some corroborative data have been obtained from measurements of the electrical ('Johnson') noise in a resistor, and at the National Physical Laboratory, J. E. Martin, T. J. Quinn and B. Chu (*Metrologia* **25**, 107–112; 1988), have used a total radiation calorimeter relying on the Stefan-Boltzmann  $T^4$  law to produce new data between –130 °C and +100 °C. This independent method confirms the results of gas thermometry carried out at the National Bureau of Standards (NBS; now the National Institute of Standards and Technology) near Washington DC, which was the first to suggest that the 1968 scale was substantially in error (L. A. Guildner & R. E. Edsinger *J. Res. natn. Bur. Stand.* **A80**, 703–738; 1976). For example, these experiments indicate a value for the 'steam point' of 99.975 °C, unexpectedly (five standard deviations) different from its conventional value of 100 °C.

ITS-90 will implement many other improvements which will have more practical effect than the numerical changes. Boiling points will be replaced by the more reproducible freezing or triple points wherever possible (the steam point itself will not be included), and the mathematical formulation will be revised to reduce errors of interpolation. The range will be extended down to 0.65 K by including helium vapour pressure equations — but the millikelvin physicist will continue to plough a lonely furrow for some time to come.

Perhaps the most significant development, and the most difficult to accomplish, has been the removal of the platinum-rhodium thermocouple which is used in the 1968 scale above 630 °C. The thermocouple is responsible for the erratic errors in that range, as shown in the figure, and it is not sufficiently reproducible for present-day standards. Its replacement has required the development of platinum resistance thermometers capable of operating with good stability up to the freezing point of silver, about 962 °C. Beyond this point ITS-90 will use radiation pyrometry, which is currently specified only above the freezing point of gold, which is at a temperature of near 1,064 °C.

With the introduction of the ITS-90, we shall for the first time have a coherent scale extending from below 1 K up to the freezing point of gold and beyond, based on instruments of proven reproducibility. The ultimate goal of a practical thermodynamic thermometer remains elusive, but the new scale should serve as the basis of improved temperature standards for a good many years. □

R. L. Rusby is in the Division of Quantum Metrology, National Physics Laboratory, Teddington TW11 0LW, UK.



Approximate differences between temperatures on the 1990 and 1968 scales.



# Late extinctions of amphibians

Andrew Milner

THE amphibian order Temnospondyli was the main group of early amphibians from the middle of the Carboniferous (315 million years ago) to the end of the Triassic (215 million years ago). During this time, these animals filled many of the niches now occupied by salamanders, crocodiles and even large freshwater fish. Until just over 10 years ago, no post-Triassic temnospondyls were known and it was reasonably assumed that all had become extinct at the end of the Triassic. Consequently, they have been incorporated as components of supposed Triassic extinction events<sup>1-3</sup>. But a series of new, little-publicized discoveries<sup>4-7</sup> extends the stratigraphical range of several temnospondyl lineages beyond the Triassic (see figure) and suggests that, as at the Cretaceous-Tertiary boundary event, amphibians passed unscathed through the events which made many reptiles extinct.

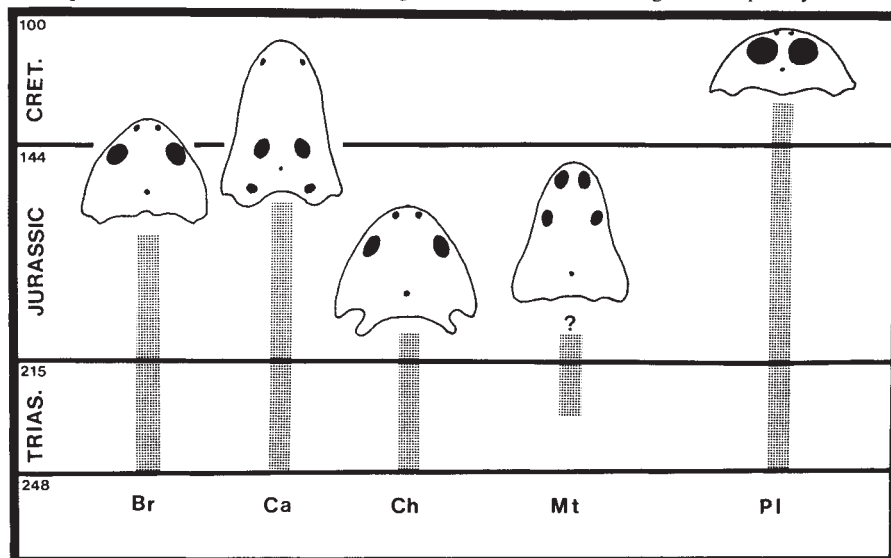
The first specimen to force a reconsideration of temnospondyl chronology was the huge complete skeleton from the Lower Jurassic (200 million years ago) of Queensland which Warren and Hutchinson<sup>4</sup> named *Siderops*. This specimen belongs to the Chigutisauridae, a family of neotenous temnospondyls (in which the larval body form is retained by the adult) previously known from the Triassic of South America and Australia. Before the significance of this single relict lineage could be determined, several more fossils were found in east and central Asia. One of these is a well-preserved skull from the Middle Jurassic (175-million-year-old) dinosaur assemblage found at Zigong, Sichuan, China<sup>5</sup> named *Sinobrachyops*, a member of the widespread family Brachyopidae. This specimen represents a different lineage of neotenous temnospondyls not previously recorded later than the Middle Triassic. Another is the series of temnospondyl bones from the later Middle Jurassic (165-million-year-old) estuarine beds of the Kirgiz region of Soviet central Asia just described by Nessov<sup>6</sup>. These bones seem to belong to the Capitosauridae, a family of crocodile-like forms. The associated fauna includes several fish more usually associated with the Triassic, together with typical Jurassic-Cretaceous reptiles. So at this locality, estuarine fish and amphibians seem to represent an entire relict ecosystem; Nessov suggests that it would have been difficult for other forms to replace families which had adapted to the osmotically demanding estuarine environment.

The most remarkable amphibian discovery is that of a temnospondyl jaw from the Lower Cretaceous (about 130-million-year-old) Strzelecki Formation of Victoria,

Australia<sup>7,8</sup>. Because the fossil was found in an assemblage of dinosaur bones it was briefly identified in an exhibition catalogue<sup>9</sup> as the mandible of an ankylosaurid dinosaur! It has now been shown by Jupp and Warren<sup>7</sup> to belong to a large temnospondyl. These authors are uncertain as to its family position, but a comparison with Triassic specimens confirms that it belongs to the Plagiosauridae and that it has specific resemblances to *Plagio-*

reassessments.

An example of this is the Triassic-Jurassic boundary region in western North America which is still in a state of stratigraphic flux and may yet force further reconsideration of the timing of temnospondyl extinctions. Murry has recently noted<sup>10</sup> an assemblage of vertebrates from a Dockum Formation horizon in Apache Canyon, New Mexico, in which the metoposaurid temnospondyl *Anaschisma* (supposedly restricted to the Upper Triassic) is associated with scales of the fish *Redfieldius* (previously found only in the Jurassic). The Metoposauridae is one of the last remaining temnospondyl families



Chronological distribution of some temnospondyl amphibian families during the Mesozoic. Br, Brachyopidae; Ca, Capitosauridae; Ch, Chigutisauridae; Mt, Metoposauridae; Pl, Plagiosauridae. Numbers on the left are in millions of years.

sternum from the Middle Triassic of Germany. The plagiosaurs were wide-headed, neotenous forms superficially resembling large flattened axolotls (salamanders). Within the past decade, therefore, it has become clear that up to three temnospondyl families — the Chigutisauridae, Brachyopidae and Capitosauridae — survived into the Jurassic and one, the Plagiosauridae, into the Lower Cretaceous.

In global terms, none of these families contributed to Triassic extinctions. But all these discoveries have been made either in Australia, which was at the end of an Australasian-Antarctic peninsula by that time, or in central and east Asia, which was cut off from the other northern continents by the Turgai Straits in the Middle Jurassic. A conservative interpretation might be that temnospondyls did become extinct in the late Triassic in Europe, Africa and the Americas while persisting in partial or complete isolation in Asia and Australasia, implying a regional rather than a global extinction event. But although a few Lower and Middle Jurassic tetrapod assemblages in Europe and the Americas have now been discovered, there is plenty of potential for further

still believed to have contributed to a late Triassic extinction event. It is not yet clear whether the fish represents an unprecedentedly early record or the amphibian a new late record. If the latter turns out to be the case, then there will be little remaining evidence that amphibians were involved in any Triassic global extinction event, at least at the family level. If, as seems likely, there are patterns of extinction and replacement among Triassic reptile families, it may be inadvisable to dilute them with amphibian data. □

Andrew Milner is in the Department of Biology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

1. Benton, M.J. *Spec. Pap. Palaeont.* **33**, 185-202 (1985).
2. Benton, M.J. in *The Beginning of the Age of Dinosaurs* (ed. Padian, K.) 303-320 (Cambridge University Press, 1986).
3. Olsen, P.E. & Sues, H.D. in *The Beginning of the Age of Dinosaurs* (ed. Padian, K.) 321-351 (Cambridge University Press, 1986).
4. Warren, A.A. & Hutchinson, M.N. *Phil. Trans. R. Soc. B* **303**, 1-62 (1983).
5. Dong, Z. *Vert. Palasiat.* **23**, 301-306 (1985).
6. Nessov, L.A. *Acta zool. Cracov.* **31**, 475-486 (1988).
7. Jupp, R. & Warren, A.A. *Alcheringa* **10**, 99-124 (1986).
8. Flannery, T.F. & Rich, T.H. *Aust. nat. hist.* **20**, 195-198 (1981).
9. Smith, B.J. (ed.) *Dinosaurs from China* 1-52 (National Museum of Victoria, 1982).
10. Murry, P.A. *J. Arizona-Nevada Acad. Sci.* **22**, 73-84 (1987).



# Unravelling atherosclerosis

James Scott

AN ELEVATED level of blood cholesterol is a prerequisite for the development of atherosclerosis, the disease that causes heart attacks and strokes. Cholesterol is transported in the blood in low-density lipoprotein (LDL) particles. The rate of endocytosis of these particles by liver cells determines the level of blood cholesterol. This process is largely mediated by the LDL receptor and its two ligands apolipoprotein (apo)-B100 and apo-E. A full explanation of lipoprotein metabolism in people with familial hypercholesterolaemia caused by defects of the LDL receptor gene, however, requires the existence of another lipoprotein receptor specific for apo-E<sup>1</sup>. Until now, the apo-E receptor has eluded purification and its existence has been questioned. With the publication by Herz *et al.*<sup>2</sup> of the structure of a protein with remarkable resemblance to the LDL receptor, this controversy may have been resolved.

Another receptor of importance in the atherogenic process and cholesterol metabolism, the macrophage scavenger receptor, has also recently been isolated<sup>3,4</sup>. It mediates the clearance of oxidatively damaged LDL by macrophages — and may hold the key to atherogenesis. These discoveries should result in an increased understanding of the structural interaction between apolipoproteins and their receptors; the regulation of these receptors and their role in growth, development, inflammation and atherogenesis;

and the evolutionary relationship between the blood clotting, complement and lipoprotein systems and the molecular cascades that act during development.

Lipoproteins consist of a hydrophobic core of triglyceride and cholesteryl-ester surrounded by a polar coat of phospholipid and free cholesterol. The surface is stabilized by specific proteins called apolipoproteins. The largest of these, apo-B100, is required for the secretion of very low-density lipoprotein (VLDL) by the liver. VLDL triglyceride is released by lipoprotein lipase in the blood capillaries, thereby converting VLDL to cholesterol-rich LDL. Apo-B100 is the sole protein of LDL and the ligand that delivers cholesterol to cells by the LDL receptor. Intestinally derived lipoproteins called chylomicrons are assembled around apo-B48, which lacks the carboxy-terminal half of apo-B100 necessary for clearance by the LDL receptor. Chylomicron triglyceride suffers the same fate as VLDL triglyceride and the chylomicron remnant circulates back to the liver, where its clearance is mediated by the interaction of apo-E with lipoprotein receptors.

Evidence for the existence of a distinct hepatic apo-E receptor comes from various sources<sup>1,2</sup>. Most convincingly, humans with familial hypercholesterolaemia and rabbits do not have elevated levels of chylomicron remnants or of other lipoproteins cleared by apo-E. Two low-molecular-mass apo-E binding proteins

have been suggested as candidates for the apo-E receptor, but this role has not been shown definitively. Herz *et al.*<sup>2</sup> have now isolated complementary DNA clones encoding a 4,544-residue membrane glycoprotein very similar to the LDL receptor which could be the apo-E receptor. The protein predicted from this DNA sequence is abundant in the liver, lung and brain, but the messenger RNA encoding it is found in most tissues.

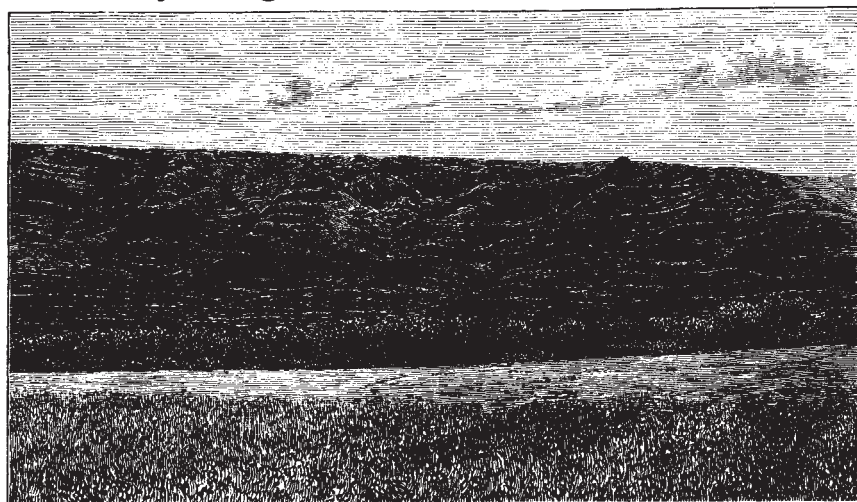
The LDL receptor is an 839-amino-acid multidomain protein<sup>2,5</sup>. The amino terminus contains seven, 40-residue, cysteine-rich repeats, each with a cluster of acidic residues near its carboxy terminus. The first repeat binds calcium which is required for ligand-receptor interaction; the other six bind regions of apo-B100 and apo-E that are enriched in basic residues.

Cysteine-rich repeats similar to those at the amino terminus of the LDL receptor are found in components of the complement complex which may stabilize cytolytic conformation of the complex by binding to domains enriched in basic residues on other proteins. Herz *et al.*<sup>2</sup> argued that proteins like the LDL receptor that interact with basic residues on apo-B100 and apo-E should contain conserved regions enriched in acidic residues, and indeed they found a complementary DNA in a mouse lymphocyte library encoding a protein very similar to the LDL receptor, which they call LDL receptor-related protein (LRP). This protein contains four repeats of the extracellular portion of the LDL receptor containing the ligand binding site, a domain similar to the epidermal growth factor (EGF) precursor, as well as a transmembrane segment and a cytoplasmic tail with two copies of the signal used by the LDL receptor for clustering in clathrin-coated pits. Although it remains to be confirmed that the protein is a lipoprotein receptor, its resemblance to the LDL receptor, together with its antibody- and calcium-binding properties, suggest that it has this function. If LRP is the apo-E receptor, the finding that its mRNA is widely expressed outside the liver implies a more general role, perhaps paralleling the involvement of apo-E in the repair of injured nerves and in the modulation of immune responses and cell growth<sup>1</sup>.

The EGF precursor, LDL receptor and LRP are all remarkably similar over a 400-amino-acid region<sup>2,5</sup>, indicating a close evolutionary relationship. In the LDL receptor, the EGF-like domain is required for acid-dependent dissociation of the ligand from the receptor in the endosome after endocytosis and for repeated recycling of the receptor to the cell surface<sup>6</sup>. This domain in LRP could have a similar role. There are six other EGF repeats not present in the LDL receptor adjacent to the plasma membrane in LRP. The EGF precursor contains a similar grouping,

100 years ago

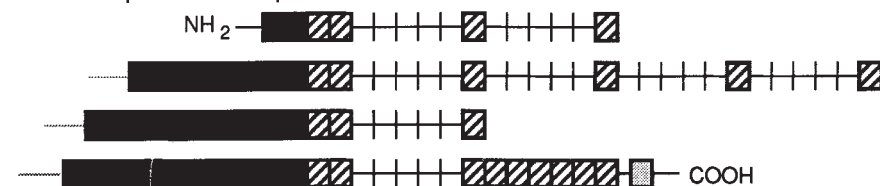
From *Nature* 39, 461; 14 March 1889.



Mr. Ernst draws attention to the formation of these ledges as observed by himself in Caracas. They are probably to be found in many places, if carefully looked for. The diagram, taken from a photograph, illustrates one of the most striking instances I know, to be found near Ballantrae, on the Ballantrae-Girvan road, Ayrshire. The ledges occur on the western face of a series of low hills, near to the sea-shore.

It seems clear that the ledges owe their origin to the action of rain-water, which would naturally penetrate below the surface covering of grass, and dissolve portions of the porous soil below. The grass layer would eventually have nothing to support it in places, and would collapse to a lower level. The effect of collapse would necessarily be a wrinkle or ledge at right angles to the ground slope. EDMUND J. MILLS.

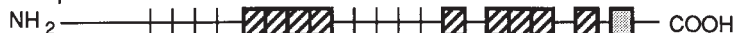
## LDL receptor related protein



## LDL receptor



## EGF precursor



Homologous domains between LRP, and LDL receptor and EGF precursor (redrawn from ref. 2). Solid boxes, homology with complement; hatched boxes, homology with EGF; grey boxes, membrane-spanning segment; vertical lines, Tyr-Trp-Thr-Asp repeats.

including the 53-residue EGF moiety<sup>1</sup>. EGF is released by proteolytic processing, and transforming growth factor- $\alpha$  (TGF- $\alpha$ ) and vaccinia virus growth factor are also released by proteolysis from membrane-bound precursors<sup>2</sup>. (Furthermore, the TGF- $\alpha$  precursor can act in its membrane-bound form without processing on the EGF receptor on adjacent cells<sup>3</sup>.) Herz *et al.* therefore suggest that a growth factor could be released from LRP by proteolytic cleavage<sup>3</sup>.

A surprisingly diverse number of other membrane-bound or secreted glycoproteins contain EGF repeats, including blood-clotting factors, proteins C, S and Z (bovine), urokinase, thrombomodulin and tissue plasminogen activator<sup>4</sup>. Proteins from the fruitfly *Drosophila* (*notch*, *delta* and *slit* loci) and from the nematode worm *Caenorhabditis elegans* (*lin-12* locus) that function in neurogenic development and cell lineage control<sup>5</sup> have tandem arrays of EGF repeats (36 in the Notch protein). The cell-attachment, chemotactic and growth-controlling properties of laminin have also been attributed to a highly conserved EGF repeat<sup>6</sup>. The available evidence suggests that the functions of these proteins in cells result from the unprocessed domain containing the EGF-like repeats<sup>9,10</sup>, which may act like the EGF-precursor homologous domain of the LDL receptor in protein-protein interactions requiring conformational change between rigid protein domains rather than by local release of growth factor.

Chemically or oxidatively modified LDL is avidly taken up by macrophages in cell cultures. It accumulates as foamy droplets in the cytoplasm, because the macrophage is unable to dispose of excess cholesterol<sup>11</sup>. The trapping of foam cells in the innermost layer of the artery wall with the release of growth factors and chemo-attractants is thought to initiate and perpetuate proliferation of smooth muscle cells and deposition of components of connective tissue that form the atherosclerotic plaque<sup>12</sup>. Endothelial cells which

line the arterial lumen can oxidize LDL<sup>13-15</sup>, and modified LDL has recently been isolated from human blood. Oxidation of LDL causes lipid peroxidation and destruction of the basic residues on apo-B100 needed for LDL receptor-mediated clearance. Modified LDL is therefore likely to be cleared by the macrophage scavenger receptor, with subsequent macrophage activation and migration into the arterial intima. The scavenger receptor has now

## ASTROPHYSICS

## Merging groups of galaxies

François Schweizer

MOUNTING evidence suggests that galaxy formation did not end with the rapid collapse of protogalactic gas clouds shortly after the Big Bang. Instead, it is becoming increasingly clear that gravitational interactions, collisions and even mergers continue to reshape and feed galaxies right to the present day. Models of pairs of colliding<sup>1</sup> and merging galaxies have recently been refined<sup>2,3</sup> to include the effects of massive halos of dark matter that are thought to surround most galaxies. On page 123 of this issue<sup>4</sup>, Joshua E. Barnes goes one step further, reporting a numerical simulation of a whole compact group consisting of six disk galaxies, each with its own halo. This work provides a fascinating glimpse of the jostling that goes on in dense groups and that leads via multiple mergers to the rapid formation of a single remnant galaxy. Interestingly, this final remnant as well as all the intermediate merger products have the characteristics of elliptical galaxies, although their basic ingredients were disks.

Barnes's paper sheds new light on two puzzles of extragalactic astronomy. First, why did galaxies form in their well-known dichotomy of shapes? (Why are there blob-like ellipticals as well as disk-like spirals?) And second, if most galaxies have extended dark halos, why have the

been purified<sup>3,4</sup>: it is a glycoprotein of relative molecular mass 260,000 with three disulphide-linked subunits of mass approximately 70,000. It is dramatically induced in macrophages that have been activated by phorbol esters. Induction of the scavenger receptor in the activated macrophage may commit the cell to engorge more modified LDL and to its destiny as an atherogenic foam cell. □

James Scott is head of the Division of Molecular Medicine, MRC Clinical Research Centre, Watford Rd, Harrow, Middlesex HA1 3UJ, UK.

1. Mahley, R.W. *Science* **240**, 622-630 (1988).
2. Herz, J. *et al.* *EMBO J.* **7**, 4119-4127 (1988).
3. Kelley, J.L., Kerbacher, J.J. & Schwartz, C.J. *Circulation* **78**, suppl. II, Abstr. 0051 (1988).
4. Kodama, T., Reddy, P., Kishimoto, C. & Krieger, M. *Circulation* **78**, suppl. II, Abstr. 0052 (1988).
5. Brown, M.S. & Goldstein, J.L. *Science* **232**, 34-47 (1988).
6. Davis, C.G. *et al.* *Nature* **326**, 760-765 (1987).
7. Scott, J. *et al.* *Science* **221**, 236-240 (1983).
8. Wong, S.T. *et al.* *Cell* **56**, 495-506 (1989).
9. Rothberg, J.M., Hartley, D.A., Walthers, Z. & Artavanis-Tsakonas, S. *Cell* **55**, 1047-1059 (1988).
10. Panayotou, G., End, P., Aumailley, M., Timpl, R. & Engel, J. *Cell* **56**, 93-101 (1989).
11. Brown, M.S. & Goldstein, J.L. *Nature* **316**, 680-681 (1985).
12. Ross, R. *New Engl. J. Med.* **314**, 488-500 (1986).
13. Parthasarathy, S., Fong, L.G., Otero, D. & Steinberg, D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 537-540 (1987).
14. Steinbrecher, U.P. *J. biol. Chem.* **262**, 3603-3608 (1987).
15. Avogaro, P., Bon, G.B. & Cazzolato, G. *Arteriosclerosis* **8**, 79-87 (1988).

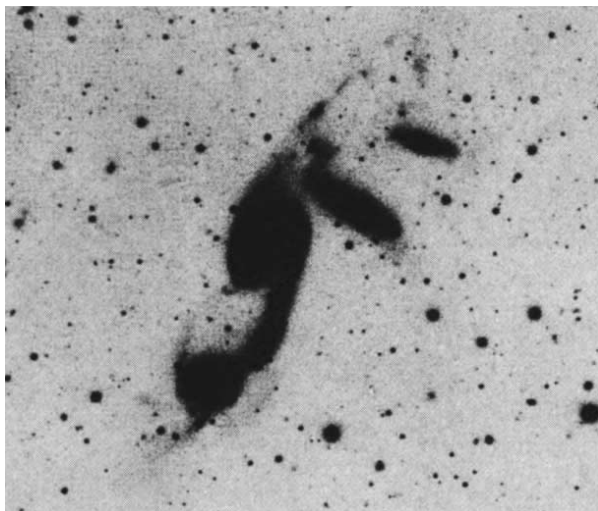
members of existing compact groups not long since collided and merged?

The classical picture of the collapse of protogalactic gas clouds suggests vaguely that differences in the initial rate of star formation might determine the final shape of a galaxy (see the recent News and Views article by M. G. Edmunds<sup>5</sup>). Yet the alternative hypothesis<sup>1</sup> that ellipticals could be remnants of merged disk galaxies, though widely disregarded at first, has recently been attracting much support. According to this hypothesis, it is disk galaxies that are formed primarily from protogalactic gas clouds. As remnants of merged disk galaxies, ellipticals are then only secondary products of galaxy formation.

A host of observations<sup>6,7</sup> supports this hypothesis, and improved model simulations of disk mergers<sup>3</sup> do, indeed, predict slowly rotating, triaxial remnants with the large internal random motions and density distributions characteristic of elliptical galaxies. A stride of sorts has been the addition of dark halos to the model galaxies, facilitated by new, fast computer codes that allow simulations with up to  $10^5$  mass particles. As it turns out, halos are remarkably efficient at soaking up orbital energy and angular momentum, thus eliminating the objection that previous



Klemola 30, a compact group of four disk galaxies in the southern skies. Gravitational interactions, and perhaps even a recent merger, have generated spiral arms and tidal tails. Luminous matter, presumably stripped from individual galaxies, pervades the group. Computer simulations of such compact groups suggest that all disks will eventually merge into one single remnant resembling a giant elliptical galaxy. Compare this photograph with Fig. 1b of Barnes's paper on page 124 of this issue. (Courtesy of John A. Graham.)



models left the merger remnants rotating too fast.

Most galaxies occur in aggregates ranging from simple pairs through groups to large clusters and superclusters with thousands of members. Among these aggregates, some compact groups of typically 4–6 galaxies separated by only a few galactic radii have been known for several decades (see figure), but interest in these increased dramatically with the publication of two catalogues<sup>8,9</sup> totalling over 100 such groups. It was soon realized that the putative dark halos surrounding individual galaxies should lead to large cross-sections and rapid merging, given that galaxies cross their groups typically in  $10^8$ – $10^9$  years. Why, then, do we still observe compact groups  $15 \times 10^9$  years after the Big Bang?

The answer probably lies in the general expansion of the Universe. As Kahn and Woltjer<sup>10</sup> pointed out long ago for our own Local Group of galaxies, the fact that the Milky Way and the Andromeda Galaxy are now approaching each other at 120 kilometres per second suggests strongly that they and other Local Group members are falling back together for the first time since the initial expansion. Similarly, as Barnes argues in this issue<sup>4</sup>, compact groups of galaxies must condense continually from the reservoir of loose groups, which are 50 times more common, through delayed infall helped, of course, by the fierce dynamical friction experienced when members occasionally collide. His simulation indicates that, if the larger two of the six disks were scaled up to the size of the Milky Way, the first two mergers in his compact group would occur within  $5 \times 10^8$  years and the whole group would have merged into one giant elliptical after only  $4 \times 10^9$  years. In this scheme, then, present-day compact groups are transient configurations of galaxies in the process of merging, or just relatively late stages of collapses delayed by the initial expansion.

A quick check of recent observations shows no contradictions with the scheme.

On the contrary, the luminosities of galaxies in Hickson groups<sup>11</sup> do, indeed, add up to total luminosities typical of giant ellipticals, the predicted end product of group evolution. Also, measured colours of group ellipticals<sup>11</sup> indicate an unusually high percentage of blue ellipticals, pointing to recent collisions that formed masses of young blue stars. Perhaps most telling is spectroscopic evidence for kinematic perturbations recently found by V.C. Rubin, D.A. Hunter and W.K. Ford (in preparation); eight out of nine brightest galaxies in Hickson<sup>9</sup> groups show severely perturbed rotation patterns suggestive of mergers, whereas only about one-third of the lesser group members show similar patterns.

Compact groups are thus emerging as the third believable source of delayed galaxy formation and transformation, after the disk-disk mergers reviewed by Toomre<sup>12</sup> and the many ongoing mergers of gas-rich galaxies revealed by the Infrared Astronomical Satellite (IRAS) about 5 years ago. Mergers, once perceived as rare oddities, now seem to be ubiquitous. Classifying galaxy morphologies 60 years ago, Hubble could scarcely have guessed how fleeting the shapes of galaxies may be. Today, the special thrill provided by Barnes's numerical experiment is to make us bystanders in a cosmic game of Russian roulette; yet instead of death from hits, there is rebirth and growth. □

François Schweizer is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA.

1. Toomre, A. & Toomre, J. *Astrophys. J.* **178**, 623–666 (1972).
2. Negroponte, J. & White, S.D. *Mon. Not. R. astr. Soc.* **205**, 1009–1029 (1983).
3. Barnes, J.E. *Astrophys. J.* **331**, 699–717 (1988).
4. Barnes, J.E. *Nature* **338**, 123–126 (1989).
5. Edmunds, M.G. *Nature* **337**, 600–601 (1989).
6. Schweizer, F. *Science* **231**, 227–234 (1986).
7. Bender, R. *Astr. Astrophys.* **202**, L5–L8 (1988).
8. Rose, J.A. *Astrophys. J.* **211**, 311–318 (1977).
9. Hickson, P. *Astrophys. J.* **255**, 382–391 (1982).
10. Kahn, F.D. & Woltjer, L. *Astrophys. J.* **130**, 705–717 (1959).
11. Hickson, P., Kintl, E. & Auman, J.R. *Astrophys. J. Suppl.* (in the press).
12. Toomre, A. in *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B.M. & Larson, R.B.) 401–416 (Yale University Observatory, New Haven, 1977).

## Water everywhere

LAST week Daedalus invented Wet Paint®, a cunning self-cleaning paint. It is mainly a hygroscopic polymer resin, which absorbs water vapour from the night air while the temperature is low and the humidity is high. But in the morning, when the paint warms up again, dispersed waxy globules in the paint melt, expanding strongly as they do so. The strong internal pressure they create in the water-swollen polymer forces water out of it by reverse osmosis. So dirt on the surface is lifted off and washed away daily. In effect, the paint acts as an automatic water-condensing dehumidifier, powered by the temperature difference between day and night.

But Wet Paint is far more than a clever way of keeping cars clean and lifting graffiti. Daedalus recalls the many areas of the world, from California to Libya, where water has to be expensively piped from hundreds of miles away, or mined from shrinking and increasingly contaminated ground-water reserves. The atmosphere contains about 1 per cent by weight of water, freely available to anyone with the wit to extract it. Wet Paint now provides that wit.

A square metre of Wet Paint should capture every day about a litre of water. So ten square metres or so could keep a man alive and tolerably clean. A hundred square metres would permit him almost an urban lifestyle. DREADCO's engineers are already devising Wet-Painted water towers whose extensive surfaces accumulate water efficiently from the night breezes, and deliver it automatically, filtered from airborne dust and dirt, every morning. They will need no power and consume no resources. At last every home and rural village can be aqueously self-sufficient.

A small-scale, distributed water source like Wet Paint will never meet the needs of heavy industry. Steel works and chemical plants will still need to be sited near lakes or rivers. But some agriculture could benefit. A desert would have to be covered several times over with Wet Paint to gain the water equivalent of normal temperate rainfall. But a Wet-Painted greenhouse, trapping and recycling the water naturally transpired by the plants within, could well flourish in such a desert. It could help an isolated and parched community to provide its own food as well as its own water.

On a small scale too, Wet Paint water condensers will find many uses. With their aid, cats and office plants and laboratory rats will be regularly supplied with water, batteries and aquaria will be topped up, sandwiches kept moist. Campers and explorers will bless the reliability and purity of their product; and paranoid tourists, warned not to drink the local water, will at last have an alternative to whisky.

David Jones

# Rhino protection in Africa

SIR—The Commentary by Leader-Williams and Albon (*Nature* 336, 533; 1988), on the allocation of resources for conservation of African big game, leads me to enquire whether any consideration has been given to dehorning of wild rhinoceroses.

For the past 30 years a dehorned male rhinoceros has lived in the National Zoological Park in Washington, DC. Dehorned animals may be handicapped in ritual sparring, and the attractiveness of rhinos for tourists might be marginally reduced (although countered by raised awareness of the poaching problem), but elimination of the rationale for rhino slaughter seems to be a simple and inexpensive measure that might save the species.

JOHN BUCK

4505 Saul Road,  
Kensington, Maryland 20895, USA

SIR—The Commentary by Leader-Williams and Albon makes disheartening reading. Sadly, I believe that the prospects may be even gloomier than they suggest.

The authors claim a linear relationship between money spent per square kilometre of habitat and the health of the populations of black rhinos and elephants in Africa. But there is an alternative interpretation of the data presented in their Fig. 4, revolving about the point representing Kenya. My explanation assumes that the Kenyan datum is reliable, an assumption that rests on the fact that the Kenyan national park system is one of the best-run in Africa.

The relationship portrayed is probably not one of linear increase, but is instead 'flat' through levels of spending represented by Kenya (say, less than \$200 per km<sup>2</sup>) and only begins to show returns when spending levels increase well into the \$200–300 per km<sup>2</sup> range. This is to say that considerable sums must be spent before returns can be expected, and at lower levels there is no reason to expect even minor success. The authors' citation of the relationship between spending and elephant populations (in the caption to Fig. 4) provides a substantially smaller  $r^2$  value, indicating even less clarity in that relationship.

I see no conflict between this alternative explanation and the data represented in Fig. 3, which show a linear relationship between effort—as log<sub>e</sub> (total patrol days/size of area/frequency of rhino sightings in 1980)—and the instantaneous rate of change in black rhino populations in the Luangwa Valley. Restricting analysis to a single area offers an important control in a continent as politically and socially diverse as Africa. The continental relationship may likewise be linear, but the data do not yet support that interpretation

to the exclusion of others.

The expectation of disappointment at lower spending levels would take some of the bite out of the frustration, and perhaps fuel the argument for greater expenditure on conservation. We are all in for a long haul, and it is high time that both poor and especially rich nations realized that the fight to save African wildlife will continue to be difficult, frustrating and costly.

RANDALL BREITWISCH

Department of Biology,  
University of Dayton,  
300 College Park,  
Dayton, Ohio 45469, USA

LEADER-WILLIAMS REPLIES—The dehorning of rhinos has been discussed by conservationists since the 1950s, but has not yet been attempted as a means of preventing poaching. This would require considerable resources, because wild rhinos first have to be caught. The Zimbabwe capture unit, for example, is working to capacity in the Zambezi Valley to remove black rhinos to safer parts of the country and catches around 75 animals per year. Nor has it been established whether dehorning could be permanent, for rhinos which lose a whole horn regrow a new one within 3 years.

Dehorning may also be counterproductive. In the short term, a poacher sighting a rhino in thick bush could shoot before finding his quarry is hornless. In the long term, mother rhinos may be less able to protect their calves from predators, such as lions or hyaenas, thereby reducing recruitment. But given the desperate plight of black rhinos, it may now be appropriate to develop a permanent method of dehorning and at least to monitor the survival of one experimental group of dehorned and control rhinos.

On the basis of one outlying point, Breitwisch argues that the relationship between spending and success at protecting

rhinos in different countries may suggest a threshold. In fact, a semilog plot does not provide any better fit to the data for rhinos ( $r^2 = 0.64$ ) than the simple regression in our Fig. 4. Furthermore, the regression for elephants (not displayed in our article) had outlying points on both lower and upper sides of the line, as expected in such a politically and socially diverse continent as Africa.

More importantly, there is little disagreement in terms of policy. If resources are scarce, we suggested that they must be concentrated at appropriate levels within small areas to achieve given objectives, in this case to prevent a decline of rhinos. The donation to Zambia, which resulted in the linear relationships for Luangwa Valley in our Fig. 3, achieved little because it was insufficiently concentrated and rhinos declined in all areas at rates far in excess of recruitment. Hence, we did not argue for 'dilute' spending because funds used in this way are, ultimately, wasted. Whether the appropriate sum per unit area is derived from an intercept or from Breitwisch's proposed threshold appears to be immaterial.

We also noted that governments in developing countries and charitable conservation agencies do not have sufficient funds to maintain large conservation areas. It may be less disheartening in future if bilateral aid organizations become increasingly involved in conservation for development. At present, only a small fraction of their annual budgets of around \$40,000 million are spent on conservation. Even a further 5–10 per cent of their budgets, harnessed into well-directed schemes requested by host governments, would greatly help to stem the erosion of the natural resource base of developing countries.

N. LEADER-WILLIAMS

Large Animal Research Group,  
Department of Zoology,  
34A Storey's Way,  
Cambridge CB3 0DT, UK

## A new planet around SN1987A

SIR—The discovery in SN1987A of a 0.508-ms pulsar, with a sinusoidal  $3 \times 10^{-3}$  Hz (peak-to-peak) modulation of 8-hour period, has been reported<sup>1</sup>. The origin of this modulation is intriguing.

A real modulation in the rotation period of the pulsar by pulsations with an 8-hour period can be excluded. The modes of pulsation of degenerate white dwarfs are relatively well understood<sup>2</sup>. The upper limit to the period of the long-period gravity modes is calculated to be around 1,000 seconds for white dwarfs, in agreement with the observations. For a fixed mass the maximum period is directly proportional to the radius of the star. A neutron star has about the same mass as a white dwarf but is three orders of magni-

tude smaller, so the maximum pulsation period will be around 1 second.

A natural explanation for the modulation is that it is due to a body orbiting the neutron star. The semi-major axis of the orbit, around a neutron star of one solar mass ( $M_{\odot}$ ), of an object with an 8-hour period is 2 solar radii. This implies a mean orbital velocity of such an object of about  $300 \text{ km s}^{-1}$ . The velocity of the pulsar of  $0.23 \text{ km s}^{-1}$ , deduced from the Doppler shift, is only a lower limit to the true velocity, because the inclination is unknown. This gives a lower limit to the mass of the orbiting planet of  $7.7 \times 10^{-4} M_{\odot}$ ; the most likely mass is about twice this (that is, about the mass of Jupiter).

The planet cannot have existed before



the supernova explosion because its present orbit would then have been inside the blue giant progenitor star. Even a planet orbiting just above the surface of the star would have been expelled from the system by the explosion because this abruptly reduced the central mass by a factor of 20. The planet must therefore be composed of matter that came from deep inside the star and went into orbit after the supernova explosion.

When SN1987A exploded, about  $18 M_{\odot}$  of matter on the outside was expelled, but the matter in the interior collapsed inwards to form the neutron star. A model<sup>3</sup> suggests that the radial velocity of matter fell below  $500 \text{ km s}^{-1}$  close to the centre of the envelope. It is conceivable that  $1.5 \times 10^{-3} M_{\odot}$  of this inner envelope may have gone into orbit around the central core assisted by rotation of the core and a slight departure from spherical symmetry in the explosion.

The matter deep inside SN1987A will consist almost entirely of heavy metals

which, within a year of the outburst, will have condensed into grains. Computer simulations of the formation of planets in our Solar System from coalescence of grains<sup>4</sup> show that the process is very rapid, taking only 100–1,000 orbital periods. For the putative SN1987A planet this corresponds to less than a year. The simulations also show that the most likely result of such coalescence (especially for more eccentric initial orbits of the dust cloud) is a single large planet, as may be the case around SN1987A.

C. MARTIN GASKELL

*Astronomy Department,  
Dennison Building,  
University of Michigan,  
Ann Arbor,  
Michigan 48109-1090, USA*

1. Middledich, J. et al. *IAU Circ. No. 4735* (1989).
2. Hansen, C. J., Winget, D. E. & Kawaler, S. D. *Astrophys. J.* **297**, 544–547 (1985).
3. Woosley, S. E. in *Supernova 1987A in the Large Magellanic Cloud* (eds Kafatos, M. & Michalitsianos, A.) 289–299 (Cambridge Univ. Press, 1988).
4. Isaacman, R. & Sagan, C. *Icarus* **31**, 510–533 (1977).

## New ways with gravitational waves

SIR—John Maddox (*Nature* **336**, 513; 1988) makes the statement that “solidly informed opinion has moved away from the instrument designs of a decade ago” and implies that researchers have shifted their effort to laser interferometer detectors. This is incorrect. Four US groups and others in Italy, Australia and China are actively engaged in research to operate and develop improved resonant bar gravitational-wave detectors.

Resonant bars remain the most sensitive gravitational-wave detectors yet developed and are the only instruments so far used to mount a serious search for gravitational radiation. There has been tremendous development and refinement of the basic design used by Joe Weber two decades ago and the potential remains for vast improvements.

Maddox's statement that resonant detectors are “out of fashion” for mechanical reasons is also completely off the mark. Although larger detectors would give proportionally larger signals, the potential sensitivity of a particular detector also depends on the methods available for motion detection. Maddox states that interferometry is the preferred method for detecting the vibrations of a bar detector but electromechanical transducer schemes, which function by the modulation of an impedance in a superconducting circuit, have been found to be superior. There is no reason to attempt to construct larger bar detectors until the full potential of existing bars has been exhausted by pushing the transducer sensitivity to the maximum. Thus the argument that the larger bars which are needed will be difficult to suspend in a vibration-free manner is specious.

Nor is the issue of the quantum limit of bar detectors an immediate concern. There is a factor of about 1,000 to go from the present bar-detector strain sensitivity level of  $3 \times 10^{-18}$  to the ‘quantum limit’. It should even be possible to surpass this limit by using a specially designed transducer to force the bar into a ‘squeezed state’, analogous to an optical squeezed state, and integrating the existing technologies for low acoustic loss materials and low electrical loss superconducting circuits. In this way, a sensitivity level of  $10^{-22}$  could be reached with bar detectors.

MARK F. BOCKO

*Department of Electrical Engineering,  
University of Rochester,  
Rochester, New York 14627, USA*

SIR—We fully share the concern (*Nature* **336**, 513; 1988) that the US National Science Foundation may turn down the proposal for a new gravity-wave detector based on laser interferometer and we are equally concerned about the financing of the equivalent European projects. We are convinced, however, that resonant bar detectors have a much greater future than is suggested, and wish to correct some statements about them.

First, bar detectors are not, of course, “suspended so as to pick up mechanical movements from the surroundings” but are substantially isolated by stacks of mechanical filters. In present detectors the external container can be hit with no effect on the output data, so it is difficult to believe that “passing traffic is a nuisance”. In an interferometer at least three mirrors are suspended and protected by mechanical filters.

Second, the measurement of the displacement of the free surface of the

suspended object is not “usually done by cementing a mirror on the surface and setting up an interferometer”. This technique has been tested without much success, but the transducers used in all the antennas that have been operating and taking data have nothing to do with interferometers.

Finally, the uncertainty principle and the corresponding quantum noise limit the performances of all measuring devices, not only of some of them, and the frequency spectrum of a bar detector is not “a complicated function of its geometry”: the longitudinal mode, which is what matters because it is best coupled to the gravitational radiation field, is separated from the other resonances by several hundred hertz.

Most probably, both types of detectors, bars and interferometers, will be pushed to the sensitivity necessary for providing useful astrophysical data. Only their use in observational work will allow an adequate comparison, because computations on paper can be misleading. We tend to believe that both will find favour, but for different classes of events.

E. AMALDI

G. V. PALLOTTINO

G. PIZZELLA

*Dipartimento di Fisica,  
Università degli Studi di Roma  
“La Sapienza”, I-00185 Roma, Italia*

## Fear of flying

SIR—Robert Parker (*Nature* **336**, 719; 1988) purports to demonstrate that a Boeing 747 crash could put 250,000 people at risk from the combustion of the depleted uranium counterweight in the tail.

His figures show that on the worst-case assumptions, 8.5 kg of uranium would reach the atmosphere in respirable form, and that if 250,000 people inhale exactly equal shares of a total of 8 kg which then concentrates on one kidney in each individual, that kidney reaches its maximum permissible uranium oxide concentration.

May one enquire how all the contaminated air gets uniformly inhaled, why the absorbed material seeks out one kidney, and why the maximum permissible concentration puts people at risk?

D. E. J. THORNTON

*9 Burford Lane,  
Lymm, Cheshire WA13 0SE, UK*

### Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

# Evolution of compact groups and the formation of elliptical galaxies

Joshua E. Barnes

Institute for Advanced Study, Princeton, New Jersey 08540, USA

The dynamical evolution of gravitationally bound groups of galaxies should account for some of the objects seen in the Universe. A numerical simulation of the merging of a compact group shows that the process may result in the formation of bright elliptical galaxies. If compact groups are interpreted as transient phases in the dynamical evolution of looser groups, the merger remnants they leave behind exemplify galaxy formation at the present epoch.

BEYOND the rich clusters of hundreds or thousands of galaxies are many smaller groups containing only a few galaxies each. Most of the nearby groups are relatively loose systems<sup>1-3</sup>, with crossing times of 1 to 10 Gyr, a significant fraction of the age of the Universe. (A crossing time is the time taken for a typical galaxy to cross the system.) Our own Local Group, for example, is an extended system which is only now collapsing out of the Hubble flow<sup>4</sup>. Deeper surveys<sup>5,6</sup> occasionally reveal much more compact groups, with estimated crossing times as short as 0.1 Gyr. These rare systems contain a rather large number of tidally disturbed galaxies<sup>5-8</sup>.

The dynamical evolution of a compact, gravitationally bound group of galaxies is an interesting problem. Theoretical arguments<sup>9</sup> indicate that such groups should undergo extensive restructuring in a few crossing times as a result of dynamical friction<sup>10</sup>, tidal disruption and galactic collisions. Numerical simulations have qualitatively confirmed this expectation<sup>11-14</sup>. These models, however, used rather unrealistic representations of the actual spatial and dynamical structure of a group of galaxies. Progress in computer and software design means that this failing can now be partly remedied.

## The numerical experiment

To illustrate what might happen in an admittedly extreme—but not inconceivable—situation, I have carried out a simulation on what is initially a moderately compact group of six disk galaxies. The initial conditions, shown in Fig. 1a, were generated using a simple binary hierarchy<sup>15-17</sup>. (In the following, experimental values are given in an arbitrary system of units in which the gravitational constant  $G = 1$ .) This sextet comprised two triplets, each of mass  $m_1 = 5$ , orbiting each other at a centre-of-mass separation  $d_1 = 3.2$ . Each triplet consisted of a single large galaxy of mass  $m_2 = 2.5$  and a pair of smaller galaxies with the same total mass, orbiting with separation  $d_2 = 2$ . Each of these pairs was composed of two galaxies of mass  $m_3 = 1.25$ , with separation  $d_3 = 1.25$ . At every level  $l$  of this hierarchy, the relative orbital speed was set at  $v_l = (2m_l/d_l)^{1/2}$  to instantaneously satisfy the virial theorem; directions for position and velocity separations were chosen randomly. The entire group had a mass of 10, a binding energy of  $-21.7$  and an angular momentum of 11.3.

The galaxy models used for the initial conditions were composite bulge/disk/halo systems with a 1:4 ratio of luminous to dark matter. The four small galaxies were structurally identical to galaxy model 1 of ref. 18, with 1,024, 3,072 and 4,096 particles

in each bulge, disk and halo, respectively. The two large galaxies were similar, but scaled up by a factor of 2 in mass and particle number and a factor of 1.414 in length, so as to keep the characteristic surface brightness fixed. The large galaxies had exponential disk scale lengths of 0.12, masses of 2.5, binding energies of  $-3.9$  and angular momenta of 0.16. Roughly scaling one of the large disks to the size of our own Galaxy, the units of length, time and mass correspond to 30 kpc, 0.22 Gyr and  $1.2 \times 10^{11} M_\odot$ , respectively.

Initially, the dark mass was entirely concentrated in dark halos around individual galaxies, as shown by the circles in Fig. 1a. Some compact groups have such low mass/luminosity ratios that little additional dark matter, beyond that bound to individual galaxies, can be present<sup>19,20</sup>. But it seems likely, on general theoretical grounds, that groups formed by gravitational clustering would possess a common dark halo, which was not included in the initial conditions.

The evolution of the system was computed using a 'tree' algorithm<sup>21</sup> modified to run more efficiently on vector machines<sup>22</sup>. Setting the accuracy parameter  $\theta$  to 0.7 and using quadrupole-moment corrections, this method provides force calculations with a relative error of  $10^{-3}$ . Forces were softened using a softening length<sup>21,22</sup> of  $\epsilon = 0.015$ , roughly 0.2 times the exponential scale length of the smaller disks. Particle trajectories were calculated using a simple leap-frog algorithm with a time-step of  $\Delta t = 1/128$ . With a total of  $N = 65,536$  particles, each time-step took  $\sim 2$  min on a Cyber 205. Over the entire calculation, energy and angular momentum were conserved to roughly one part in  $10^3$ .

By time  $t = 2$ , two mergers have occurred, leaving four distinct galaxies in the system. Figure 1b shows the group at this time; the two merger remnants are characterized by their long tidal tails. The remnant at  $x = -0.8$  involves the pair of small disk galaxies initially associated with the large M81-like spiral at  $x = -2.3$ . The remnant at  $x = 2$  results from the merger of a small disk galaxy and a large one, and comprises the most massive object in the group at this stage. The dark halos surrounding the merger remnants are much more extended than those around the original galaxies, reaching roughly as far as do the luminous tails and even overlapping to some degree.

Around time  $t = 6$ , the two merger remnants themselves undergo a moderately close encounter, which further distends their halos and liberates a diffuse spray of luminous particles. Following this encounter, the more massive remnant takes up residence at the centre of the group. By time  $t = 8$ , shown in Fig. 1c, the smaller remnant is at  $x = -2.2$ , enveloped by a cloud of tidal debris. The two surviving disk galaxies are as yet relatively undisturbed; the larger, at  $x = 1.3$ , is on a roughly circular orbit about the central remnant, whereas the smaller, at  $x = 2.2$ , is falling back into the group on an approximately radial orbit.

The disk galaxies are disrupted over the next four time units. This is the outcome of a series of close encounters, the first of which is a collision between the smaller disk galaxy and the central remnant at time  $t \approx 10$ , followed by a passage of the smaller remnant at time  $t \approx 10.5$ . By this stage, however, a description of the behaviour of the system in terms of two-galaxy encounters is inadequate; interactions between three or even all four galaxies are seen. At time  $t = 12$ , shown in Fig. 1d, the larger disk galaxy is beginning to merge with the central remnant.



The smaller disk galaxy, having been stripped of much of its mass, is closely orbiting the centre; it merges at time  $t \approx 14$ . Only the small remnant, at  $y = 1.4$ , is at this time still distinct from the central relic; it finally merges at time  $t \approx 18$ .

### Merger remnants

The compact group studied here produces merger remnants spanning a range of four in mass as it evolves from six disk galaxies to a single star-pile. These remnants result from collisions between galaxies of both similar and dissimilar types, with mass ratios of up to 5:1.

Figure 2 shows surface brightness profiles, assuming equal mass-to-light ratios for bulge and disk particles, of the small remnant at  $t = 4$  and the large remnant at  $t = 14.5$ . These profiles were measured by binning luminous particles according to projected distance  $r$  from their respective galactic centres, and plotted as 'magnitudes', or  $-2.5 \log$  (particle surface density), against  $r^{1/4}$ . Both remnants exhibit de Vaucouleurs profiles—straight lines of arbitrary slope, characteristic of elliptical galaxies—from a few times the softening length out to the last points measured, a result consistent with earlier merger models<sup>18,23</sup>. The effective radii  $r_e$  of the small and large remnants are roughly 0.09 and 0.17 respectively. Note that the latter remnant has a higher surface brightness at all radii, despite the fact that both remnants were assembled from disk galaxies scaled

to have the same characteristic surface brightness. However, the surface brightness at the corresponding  $r_e$  is slightly lower for the large remnant than for the small one.

The spherically averaged density profiles of the five merger remnants produced in this experiment are quite similar to those resulting from simpler mergers between equal-mass disk galaxies<sup>18</sup>. At small radii, particles from the bulges of the original galaxies predominate, but beyond  $\sim r_e$  the dark halos contribute most of the mass. Within the uncertainties, the distribution found in these merger remnants seems consistent with what little is known about dark matter in elliptical galaxies<sup>24</sup>. The strong concentration of bulge particles at the centres of these relics shows that even multi-stage mergers can preserve significant radial population gradients<sup>25</sup>.

To study the dynamical structure of the central parts of the remnants, moment-of-inertia tensors, kinetic-energy tensors and angular-momentum vectors were computed using bulge particles within a radius of 0.5 as tracers. This cutoff excludes the few per cent of the outlying particles that wander rather far from the centres of their galaxies, which would otherwise dominate the moment of inertia and angular momentum. The eigenvectors of the inertial and kinetic tensor remain roughly parallel over most of the lifetime of each remnant, and the corresponding eigenvalues are reasonably stable, so the centres of the remnants are close to dynamical equilibrium. All five remnants are triaxial,

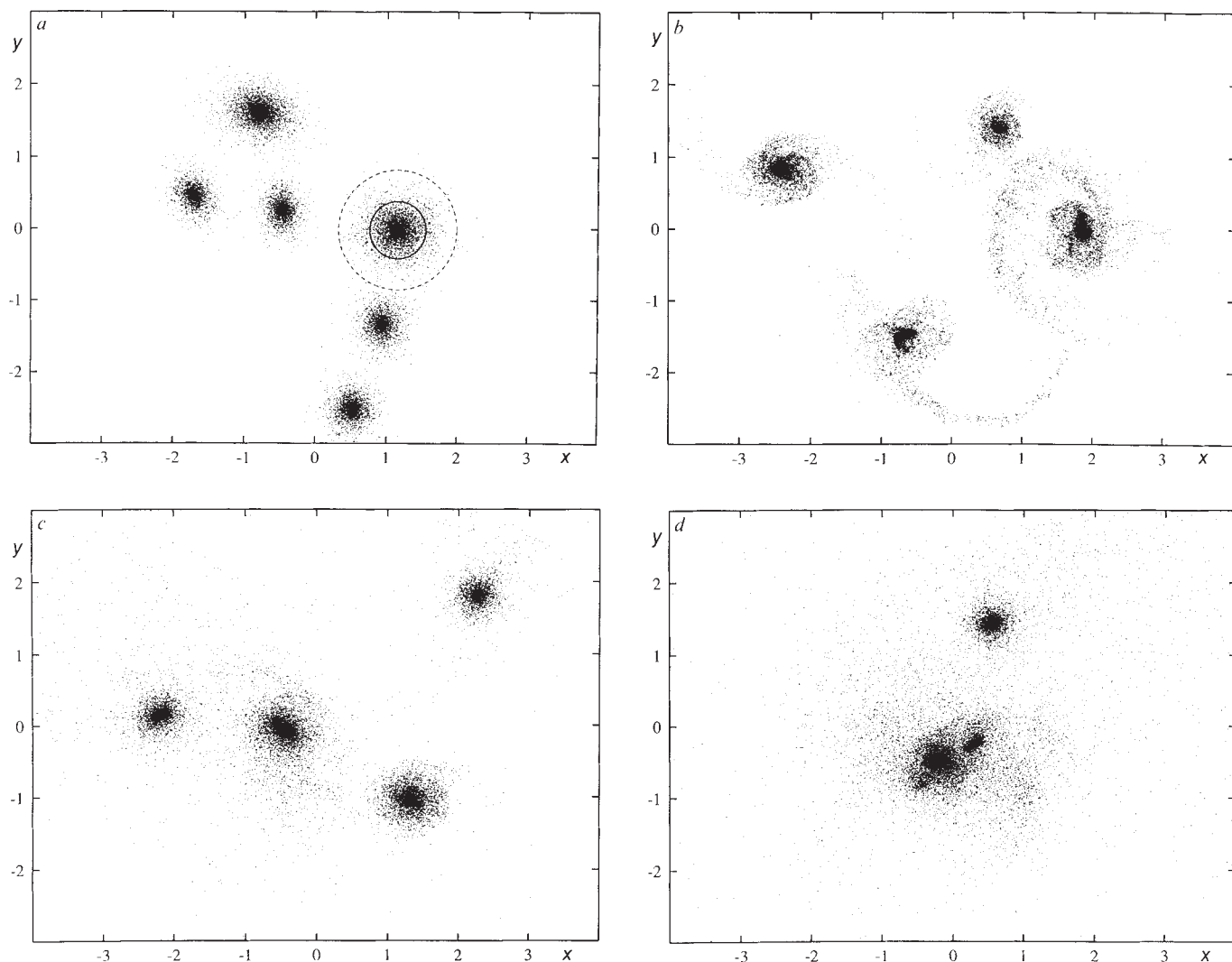


FIG. 1 *a*, Group model, initial conditions. Only the 32,768 luminous particles are plotted. Solid and dashed circles around one galaxy enclose 50% and 90% of the halo mass, respectively. *b*, Time  $t = 2$ . The two merger remnants are marked by massive tidal tails. *c*, Time  $t = 8$ . The larger merger remnant

has taken up residence at the centre of the group. *d*, Time  $t = 12$ . The larger disk galaxy has just been accreted; the smaller one is just slightly above and to the right of the central remnant.

with major-to-minor axis ratios of  $\leq 1.8$ ; prolate forms dominate but more nearly oblate ellipsoids also occur. The angular-momentum and minor-axis vectors of the remnants are all roughly parallel.

As an example, consider the larger merger remnant discussed above. This is a prolate triaxial ellipsoid, with an axial ratio of 6:7:10 inferred from the inertial tensor. Figure 3a shows a projection onto the plane containing the major and intermediate axes at time  $t=14.5$ . The velocity distribution is strongly anisotropic, with dispersion a factor  $\sim 1.4$  times higher along the major axis than along the minor axis. This anisotropy determines the shape of the remnant. The figure turns anticlockwise with a pattern speed of  $\sim 1.5$  rad per time unit, a typical star in the central ellipsoid completing many tens of orbits while the figure turns once. Much of the angular momentum responsible for this rather slow tumbling was imparted by the most recent victim, the small disk galaxy swallowed at  $t \approx 14$ .

This small disk galaxy also gives rise to interleaved shells similar to those found in some elliptical galaxies<sup>26</sup>. These shells may be seen in Fig. 3b, which shows the same projection as Fig. 3a but plots only the luminous particles from the bulge and inner third of the disk of the last victim. The shell structure is shown again in Fig. 3c in which spherical radius is plotted against radial velocity for these same particles. The four turning points between  $0.15 \leq r$  and  $r \leq 0.6$  correspond to the shells seen in Fig. 3b. That shells are found in this completely self-consistent calculation supports the formation mechanism proposed on the basis of test-particle experiments<sup>27</sup>. By time  $t=15$  the inner shells are much less well defined. Thus, the complex structure of this remnant reflects its formation in a series of mergers. Each new galaxy contributes another dynamical component to the remnant, eventually blended in by phase mixing and further violent relaxation.

## Extragalactic demographics

If compact groups rapidly evolve and merge, as these numerical experiments indicate, then how do they form, and what do they become? It is improbable that such short-lived systems appear only in the present epoch; rather, it seems likely that compact groups of galaxies have been forming and coalescing for most of the age of the Universe, and that the number of groups per comoving volume has evolved only slowly over cosmological time. These assumptions can be checked in principle by looking back to earlier epochs, and the observations<sup>28</sup> of lumpy, extended 'protogalaxies' at redshifts  $z \geq 1$  already offer some support for this suggestion.

The number density of compact groups would decline very rapidly if new ones did not form as old ones merge. Dynamical evolution of loose groups may maintain the supply of compact ones<sup>19</sup>. In loose groups the galaxies seem to be well-separated, but this apparent isolation is misleading if galaxies are surrounded by extended dark halos. Cosmological experiments<sup>29</sup> show that in such loose systems distinct dark halos spiral together and merge after only one or two group crossing times, as a result of the stickiness of extended interacting systems<sup>30</sup>. A galaxy residing within such halos would inevitably experience the same gravitational undertow as these halos feel.

If loose groups evolve into compact ones at a steady rate, the relative numbers of loose and compact groups observed should be proportional to their respective lifetimes. In both cases these lifetimes are of the order of a crossing time. These simplistic considerations suggest that the number of groups per unit volume with crossing time  $t_c$  should be roughly proportional to  $t_c$ . This relation would be modified if most loose groups decay by a series of mergers without ever evolving into compact systems.

Observations show that compact groups are rare: samples limited by both magnitude<sup>5</sup> and redshift<sup>6,31</sup> yield number densities of  $(1-2) \times 10^{-6} \text{ Mpc}^{-3}$ , assuming a value of  $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1} \approx (20 \text{ Gyr})^{-1}$  for Hubble's constant. These

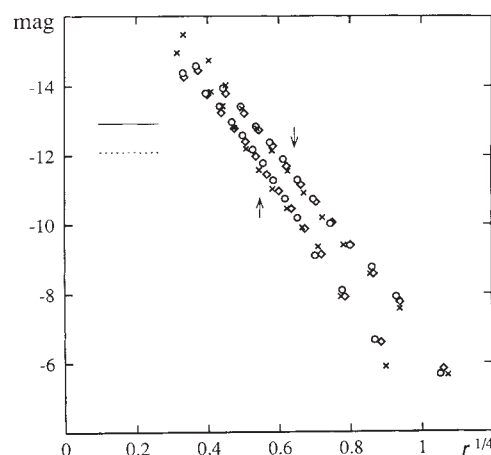


FIG. 2 Projected surface brightness profiles of two merger remnants. Three sets of symbols, corresponding to three orthogonal projections, are plotted for each remnant; upper (lower) profiles refer to large (small) remnant. Vertical arrows indicate effective radii  $r_e$ . Horizontal solid and dotted lines indicate brightness of initial bulges at half-light radii and initial disks at centre, respectively.

groups have crossing times of  $0.01-0.03 H_0^{-1}$ . Loose groups are much more common; one catalogue<sup>32</sup> derived from the CfA redshift survey<sup>33</sup> yields number-density estimates of  $5 \times 10^{-5} \text{ Mpc}^{-3}$ . Median crossing times for these loose groups are  $\sim 0.6 H_0^{-1}$ . Within the large uncertainties, the abundance ratio of compact to loose groups is consistent with the ratio of their estimated lifetimes; dynamical evolution of loose groups seems capable of producing compact groups at the required rate. The rough estimates given here do, however, imply that a significant fraction of loose groups decay via compact configurations.

The remnants left behind when compact groups merge have been accumulating in the Universe for nearly a Hubble time ( $H_0^{-1}$ ); there should be  $\sim 10^2$  group remnants for each group with a lifetime of  $0.01 H_0^{-1}$ . Taking into account the range in compact-group lifetimes, the expected number density of remnants is  $\sim 10^{-4} \text{ Mpc}^{-3}$ , which is nearly two orders of magnitude lower than the value obtained in a previous study<sup>34</sup>; the latter did not allow for the broad dispersion of group luminosities and therefore adopted a value for the number density of compact groups that was far too high (P. Hickson, personal communication).

The results presented earlier suggest that group merger remnants would be classified as elliptical galaxies. From the estimate given above, it seems possible that compact groups produce a significant fraction of all ellipticals; the characteristic number density of bright galaxies is  $\sim 2 \times 10^{-3} \text{ Mpc}^{-3}$ , of which between 10% and 20% are elliptical galaxies. It is sometimes suggested that these remnants should be isolated 'field' galaxies, but this presupposes that compact groups are themselves an isolated population. Observations show, however, that even those compact groups selected according to an isolation criterion often have nearby neighbours<sup>9,35,36</sup>. This is expected on the basis of gravitational clustering<sup>9,37</sup>, as the formation and evolution of groups is more rapid in high-density regions<sup>29</sup>. The elliptical galaxies produced by mergers in groups would be caught up by subsequent growth of larger-scale structure<sup>38,39</sup>, and are found today in regions of still higher density, where further merging is suppressed by the high velocities with which galaxies encounter each other<sup>40</sup>.

## Discussion

Starting from a hierarchy of galaxies, the compact group in these numerical simulations evolved into a system dominated by a central merger remnant, which then accreted the remaining members. The whole process took only a few orbital times,



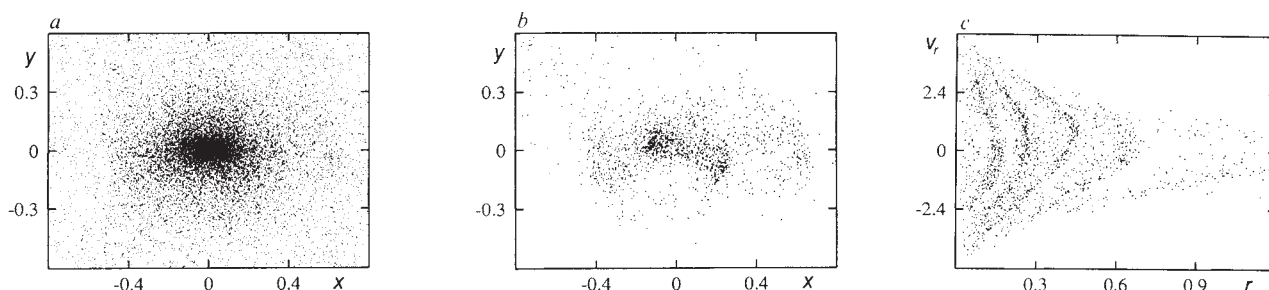


FIG. 3 *a*, Large merger remnant at  $t=14.5$ , projected onto the major-axis-intermediate-axis plane. *b*, Same as *a*, but showing only particles from the

inner part of the last galaxy accreted. *c*, Phase-space plot of the selected particles. The four turning points correspond to the shells seen in *b*.

which is consistent with earlier models<sup>11–14</sup>. Such rapid dynamical evolution seems to be a general feature of ‘lumpy’ self-gravitating systems<sup>9</sup>. Earlier numerical studies, however, gave little information about the nature of the remnants produced when compact groups decay. My calculation gives some indication of the outcome of this process: mergers in compact groups produce dynamically ordinary elliptical galaxies. This appears to be a natural outcome of the efficient braking of luminous galaxies against the halos of dark matter<sup>41</sup>, which produces remnants with the low angular momentum and characteristic luminosity profiles of bright elliptical galaxies<sup>18</sup>.

The scarcity of compact groups is unsurprising, in view of their estimated lifetimes. Indeed, these lifetimes are so short that it seems clear that the compact groups we find today must have formed relatively recently, rather than constituting the remains of a once much larger population. Dynamical evolution of loose groups can in principle supply compact groups at the required rate, according to the estimates of group number densities and lifetimes presented above. The expected number of group remnants is a significant fraction of the number of elliptical galaxies. Thus, detailed dynamical models combine with demographic arguments to give a coherent picture of group evolution and elliptical-galaxy formation.

Further numerical experiments that include loose groups and groups with large amounts of dark matter could help to test this picture. Amongst other things, such calculations might show how several galaxies in a loose group manage to come together to form a compact configuration before merging with each other. The evolutionary hypothesis advocated here predicts that a significant fraction of loose groups turn into compact ones.

Observational studies should focus on the nature of the compact groups already catalogued<sup>6,31</sup>. Careful, detailed observation

and analysis<sup>19,42</sup> of a small but diverse selection of compact groups would help to establish the characteristics of these systems. An estimate of the star formation rate in compact groups, following the approach already used for Arp galaxies<sup>43</sup>, is essential for a precise interpretation of mass-to-light ratios or number densities. Cross-correlation of compact-group catalogues with tracers of large-scale structure, and automatic recognition of compact groups on deep optical surveys, could quantify the large-scale distribution of these objects.

Finally, a theoretical framework for the interpretation of compact-group surveys is greatly needed. Samples selected homogeneously on the basis of their apparent properties<sup>5,6</sup> may nonetheless prove to be heterogeneous in their intrinsic properties, as evidenced for example by the rather flat redshift distribution of Hickson groups<sup>31</sup>. To deal with selection effects in such circumstances, theoretical models can be projected onto the observational plane, an approach already applied to loose groups<sup>32</sup>. A general treatment of the observational selection process, starting from a well-motivated phenomenological model for compact groups, would seem to be required before statistical studies of compact group catalogues can generate reliable conclusions.

Compact groups are only one of several settings in which galaxies are seen to interact with each other and in which merger remnants may be produced. For example, individual pairs of galaxies also account for a significant number of merger remnants<sup>41</sup>. In common with interacting pairs, compact groups are short-lived objects, and the relatively small number seen today nonetheless implies the production of a substantial number of elliptical galaxies over a Hubble time. Compact groups provide evidence that galaxy formation, or rather galaxy transformation, is indeed a continuing process. □

1. De Vaucouleurs, G. *Stars and Stellar Systems, Galaxies and the Universe* **9** (eds Sandage, A., Sandage, M. and Kristan, J.) 557–600 (Univ. Chicago Press, 1975).
2. Turner, E. L. & Gott, J. R. *Astrophys. J. Suppl.* **32**, 409–427 (1976).
3. Geller, M. J. & Huchra, J. P. *Astrophys. J.* **52**, 61–87 (1983).
4. Lynden-Bell, D. *The Observatory* **101**, 111–114 (1981).
5. Rose, J. A. *Astrophys. J.* **211**, 311–318 (1977).
6. Hickson, P. *Astrophys. J.* **255**, 382–391 (1982).
7. Vorontsov-Velyaminov, B. A. *Atlas and Catalog of Interacting Galaxies* (Sternberg State Astr. Inst., Moscow, 1959).
8. Arp, H. C. *Astrophys. J. Suppl.* **14**, 1–20 (1966).
9. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341–358 (1978).
10. Hickson, P., Richstone, D. O. & Turner, E. L. *Astrophys. J.* **213**, 323–326 (1977).
11. Carnevali, P., Cavaliere, A. & Santangelo, P. *Astrophys. J.* **249**, 449–461 (1981).
12. Ishizawa, T., Matsumoto, R., Tajima, T., Kageyama, H. & Sakai, H. *Publ. astr. Soc. Japan* **35**, 61–76 (1983).
13. Barnes, J. E. *Mon. Not. R. astr. Soc.* **215**, 517–536 (1985).
14. Mamon, G. A. *Astrophys. J.* **321**, 622–644 (1987).
15. Soneira, R. M. & Peebles, P. J. E. *Astr. J.* **83**, 845–861 (1978).
16. Peebles, P. J. E. *Astr. Astrophys.* **68**, 345–352 (1978).
17. White, S. D. M. & Negroponte, J. *Mon. Not. R. astr. Soc.* **201**, 401–414 (1982).
18. Barnes, J. E. *Astrophys. J.* **331**, 699–717 (1988).
19. Rose, J. A. & Graham, J. A. *Astrophys. J.* **231**, 320–326 (1979).
20. Trimble, V. A. *Rev. Astr. Astrophys.* **25**, 425–472 (1987).
21. Barnes, J. E. & Hut, P. *Nature* **324**, 446–449 (1986).
22. Barnes, J. E. *J. comp. Phys.* (submitted).
23. White, S. D. M. *Mon. Not. R. astr. Soc.* **189**, 831–852 (1979).
24. Fall, S. M. *Nearly Normal Galaxies: From the Planck Time to the Present* (ed. Faber, S. M.) 326–331 (Springer, New York, 1986).

25. White, S. D. M. *Mon. Not. R. astr. Soc.* **191**, 1P–4P (1980).
26. Malin, D. F. & Carter, D. *Astrophys. J.* **274**, 534–540 (1983).
27. Quinn, P. J. *Astrophys. J.* **279**, 596–609 (1984).
28. McCarthy, P. et al. *Astrophys. J.* **319**, L39–L44 (1987).
29. Frenk, C. S., White, S. D. M., Davis, M. & Efstathiou, G. *Astrophys. J.* **327**, 507–525 (1988).
30. Alladin, S. M. *Astrophys. J.* **141**, 768–798 (1965).
31. Hickson, P., Kindt, E. & Huchra, J. P. *Astrophys. J.* **331**, 64–70 (1988).
32. Nolthenius, R. & White, S. D. M. *Mon. Not. R. astr. Soc.* **235**, 505–530 (1987).
33. Davis, M., Huchra, J., Latham, D. W. & Tonry, J. *Astrophys. J.* **253**, 423–445 (1982).
34. Williams, B. A. & Rood, H. J. *Astrophys. J. Suppl.* **63**, 265–294 (1987).
35. Sulentic, J. W. *Astrophys. J.* **322**, 605–617 (1987).
36. Rood, H. J. & Williams, B. A. *Astrophys. J.* (in the press).
37. Layzer, D. *Stars and Stellar Systems, Galaxies and the Universe* **9** (eds Sandage, A., Sandage, M. & Kristan, J.) 665–723 (Univ. Chicago Press, 1975).
38. Jones, B. J. T. & Efstathiou, G. *Mon. Not. R. astr. Soc.* **189**, 27–38 (1979).
39. Aarseth, S. J. & Fall, S. M. *Astrophys. J.* **236**, 43–57 (1980).
40. Ostriker, J. P. *Comments Astrophys.* **8**, 177 (1980).
41. Toomre, A. *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–426 (Yale Univ. Obs., New Haven, 1977).
42. Rose, J. A. *Astrophys. J.* **231**, 10–22 (1979).
43. Larson, R. B. & Tinsley, B. M. *Astrophys. J.* **291**, 46–59 (1978).

ACKNOWLEDGEMENTS. I thank the participants of the 1988 Durham workshop, ‘The Epoch of Galaxy Formation’, for useful comments, Alar Toomre for careful readings of the manuscript, and Lars Hernquist for suggestions. I acknowledge the support of a New Jersey High Technology Grant, and thank the John von Neumann Center for computer time.

# Mutational analysis of a protein-folding pathway

David P. Goldenberg, Richard W. Frieden, Julie A. Haack & Thomas B. Morrison

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

The effects of amino-acid replacements on the disulphide-coupled folding pathway of bovine pancreatic trypsin inhibitor have been examined. Replacements at three sites destabilize the native protein relative to the unfolded state, but have different effects on the relative stabilities of the disulphide-bonded folding intermediates, thus allowing the roles of the altered residues during folding to be distinguished.

THE mechanisms by which unfolded polypeptide chains spontaneously fold into biologically active three-dimensional structures have been studied by various experimental techniques, and considerable progress has been made in identifying and characterizing intermediates in the folding reactions of several proteins<sup>1,2</sup>. Relatively little is known, however, about the conformations of these intermediates or the interactions that stabilize them, and there is not yet a consensus on the general features of folding transitions<sup>3-6</sup>. Some models of folding propose that structures in different regions of the polypeptide chain form in a specific order<sup>1,7</sup>, whereas other models suggest that the polypeptide can fold through many different pathways all of which converge on the final native structure<sup>3,8,9</sup>.

Examining the effects of changes to the amino-acid sequence on the kinetics of folding and unfolding promises to be a means of characterizing protein folding mechanisms further<sup>10-17</sup>. By comparing the effects of amino-acid substitutions at different sites on the stabilities of folding intermediates it may be possible to determine whether different residues and interactions contribute to stability and structure in a preferred order. Structural characterization of mutationally altered intermediates should provide additional details about the roles of specific residues during folding. Studies of several proteins have shown that single amino-acid replacements can not only destabilize the native state (relative to the unfolded state) by as much as 5 kcal mol<sup>-1</sup> (refs 18-21), a significant fraction of the net stability of most proteins, but can also alter the kinetics of folding reactions, both *in vitro* and *in vivo*<sup>11-17</sup>. There have been, however, few studies of mutational effects on the kinetics and equilibria of individual steps in folding, largely because there are few folding reactions for which multiple steps have been defined experimentally.

One of the best-characterized protein folding reactions is the disulphide-coupled refolding of reduced bovine pancreatic trypsin inhibitor (BPTI, Fig. 1)<sup>2,22</sup>. Native BPTI can be fully unfolded by reducing its three disulphides and efficiently refolds if the thiol-disulphide redox potential is adjusted to favour re-formation of the disulphides. Disulphide-bonded intermediates in unfolding and refolding have been chemically trapped, physically separated, and individually characterized. The conformations of the intermediates have been studied by several techniques<sup>23-25</sup>, and their relative stabilities have been determined by measuring the forward and reverse rates of each step<sup>22</sup>. The BPTI folding pathway is outlined in Fig. 1b.

Because the effects of amino-acid replacements on individual steps in BPTI folding can be examined, the protein is exceptionally well suited for a mutational analysis of the mechanism

of protein folding. In the present study, four substitutions at three sites have been examined. These substitutions all destabilize the native conformation but have different effects on the stabilities of the various intermediates, indicating that different amino-acid residues have distinct roles at different stages of folding.

## Design of mutations

To determine whether the various folding intermediates are stabilized by interactions present in the native conformation, amino-acid replacements were introduced to disrupt certain interactions present in the crystal structure of BPTI<sup>26</sup> (Fig. 1a). Site-directed mutations were made at the codons for Tyr 23, Tyr 35 and Asn 43 of the cloned BPTI gene<sup>27</sup>. These three residues are largely buried in the native conformation and seem to play important roles in stabilizing the native state. Tyr 23 forms part of the interface between the  $\alpha$ -helix and  $\beta$ -sheet in the native protein, along with Tyr 21 and the 30-51 disulphide bond. Tyr 35 is located near the opposite end of the  $\beta$ -sheet from Tyr 23, close to the 14-38 disulphide bond and the trypsin-binding site. Side-chain atoms of Asn 43 participate in three hydrogen bonds with backbone atoms. The importance of these three residues

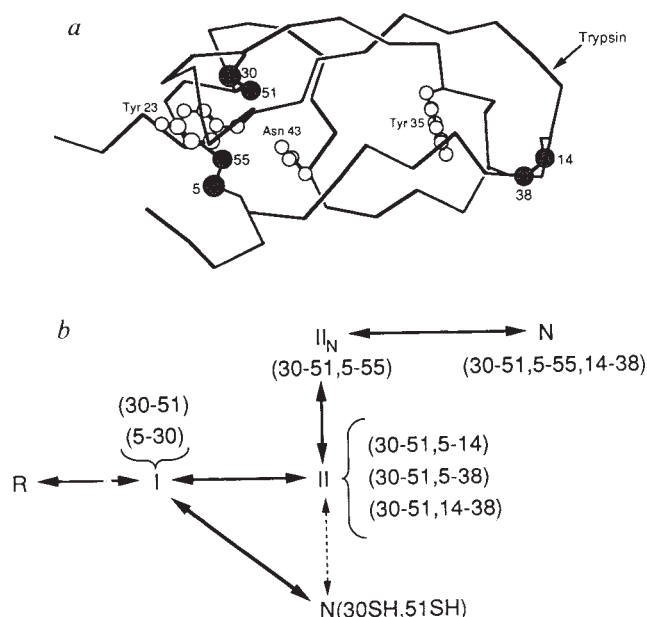


FIG. 1 Three-dimensional structure of wild-type native BPTI (a) and the disulphide-coupled folding pathway (b). The  $\alpha$ -carbon backbone of native BPTI was drawn from the Form I crystal coordinates<sup>26</sup>. Balls and sticks, side-chain atoms of the residues altered in the present study and the disulphide-bonded cysteines. Arrow, the peptide bond, between Lys 15 and Ala 16, that enters the active site of trypsin. The folding pathway is defined in terms of the disulphide bonds present in the intermediates<sup>22</sup>. R, the reduced protein; I and II, populations of intermediates with one and two disulphide bonds, respectively. II<sub>N</sub>, native-like intermediate with two of the three disulphide bonds of the native protein, N, the native protein with three disulphide bonds; N(30SH, 51SH), a native-like two-disulphide species in which the thiols of cysteines 30 and 51 are buried in the native conformation. This form can be considered an off-pathway species, as it does not readily form a third disulphide and only very slowly interconverts with the other two-disulphide intermediates.



is indicated by their high degree of conservation among proteins homologous to BPTI<sup>28</sup>.

The codons for tyrosines 23 and 35 were each altered to codons for Phe, Leu, Ala and Gly. Asn 43 was changed to Asp, Leu, Ala and Gly. These changes were chosen to disrupt interactions in the native conformation without increasing the size of the buried residues in order to minimize conformational perturbations of the native state.

The mutant BPTI genes were introduced into an *Escherichia coli* expression plasmid that directs the synthesis of BPTI as a fusion to the leader sequence of a bacterial outer membrane protein<sup>29,30</sup>. Wild-type BPTI produced in this way is processed to yield the same amino terminus as that of the mature bovine protein and folds to an active conformation in *E. coli*<sup>30,31</sup>. Although the three altered residues are evolutionarily conserved, the only mutant for which active trypsin inhibitor could not be recovered was Asn 43 → Asp, presumably because burying a charge is extremely unfavorable energetically.

Four of the mutant proteins, Tyr 23 → Leu, Tyr 35 → Gly, Asn 43 → Gly and Asn 43 → Ala were selected for further study. These

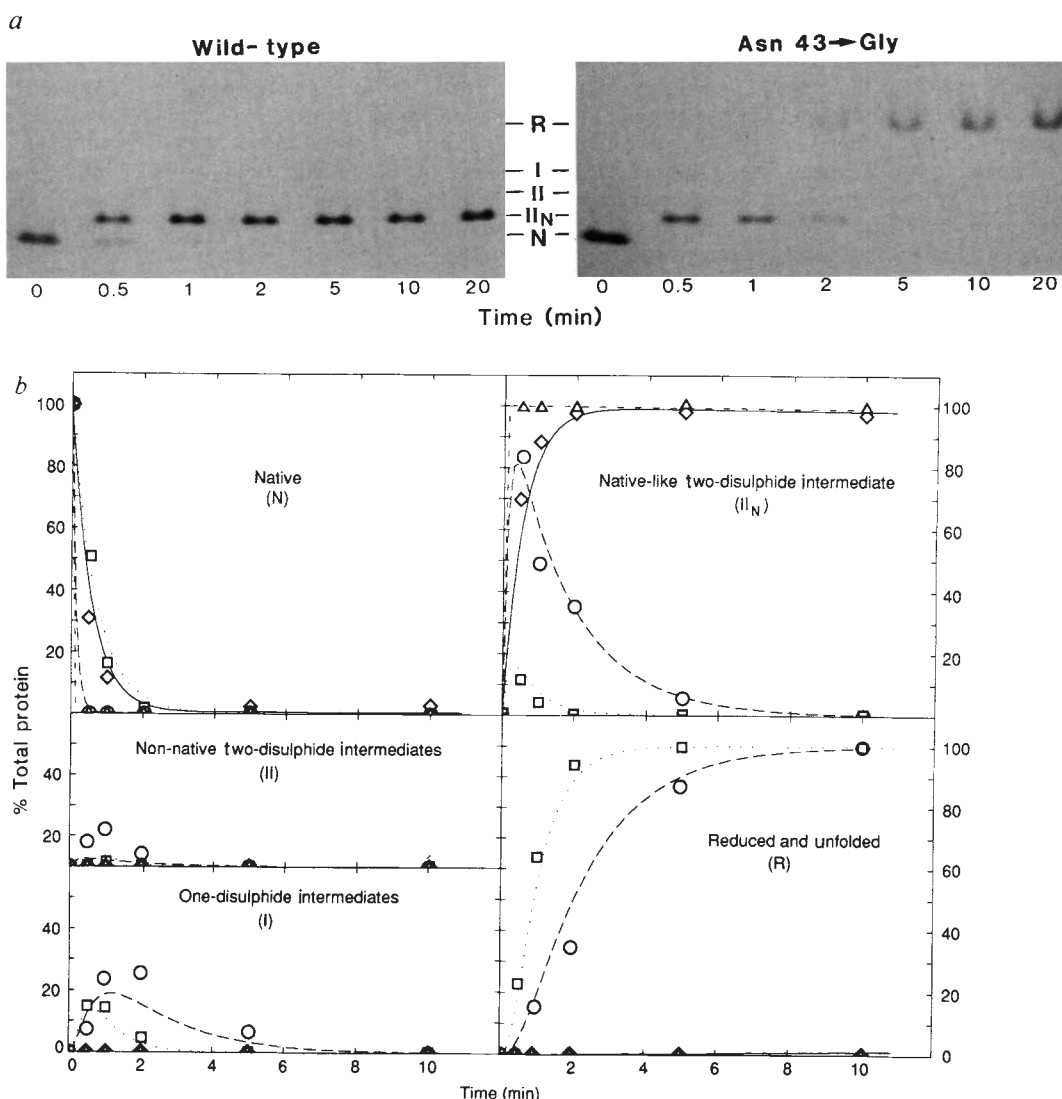
substitutions were chosen to represent the three-sites because although it seemed probable that they would destabilize the native state, they did not greatly reduce the yield of protein. The four proteins were purified and the kinetics of their folding and unfolding were measured and compared with those determined previously for the wild-type protein<sup>22</sup>.

### Kinetics of unfolding and refolding

The unfolding and refolding reactions of the mutant proteins were studied using the methods and conditions (25 °C, pH 8.7) of previous studies of the wild-type protein. Unfolding and refolding were promoted using mixtures of reduced and oxidized dithiothreitol (DTT<sub>SH</sub> and DTT<sub>S</sub>, respectively) to facilitate the breakage and formation of disulphide bonds. Under these conditions, the formation of a protein disulphide is freely reversible and its rate is proportional to the rate of the intramolecular process that brings the cysteine thiols together<sup>22</sup>. The progress of the reactions was followed by trapping partially folded states of the proteins with iodoacetic acid, a reagent that blocks free thiols. Trapped molecules with 0, 1, 2 or 3 disulphide bonds

FIG. 2 Kinetics of unfolding of wild-type and mutant forms of BPTI. *a*, Gel electrophoresis of intermediates trapped in the unfolding of wild-type and Asn 43 → Gly BPTI in the presence of 1 mM DTT<sub>SH</sub> at 25 °C, pH 8.7. Under these conditions, one of the three disulphides of native BPTI (14–38) is rapidly reduced to yield a native-like two-disulphide intermediate (II<sub>N</sub>). Further unfolding is slow, with a half-time of about 10 hours. Asn 43 → Gly, Asn 43 → Ala and Tyr 23 → Leu BPTI are much more rapidly unfolded. *b*, Quantitative analysis of the kinetics of unfolding of wild-type (diamonds, solid line), Tyr 23 → Leu (squares, dotted line), Tyr 35 → Gly (triangles, short-dashed lines) and Asn 43 → Gly (circles, long-dashed lines) BPTI. The symbols indicate the relative concentrations of the trapped native, reduced and intermediate species determined by densitometry of Coomassie blue-stained gels. The curves represent the concentrations predicted by kinetic simulations based on Equation 1 and the rate constants listed in Table 1.

**METHODS.** Synthetic oligonucleotides were used to direct mutations<sup>43</sup> at codons 23, 35 and 43 of the BPTI gene, previously cloned in M13 mp19 and fused to sequences encoding the signal sequence of a bacterial outer membrane protein, OmpA<sup>27,30</sup>. Mutant genes were fully sequenced and cloned into the expression plasmid preillac A3<sup>44</sup>. BPTI produced as a fusion to the OmpA signal sequence is correctly processed and folds in *E. coli*. Mutant and wild-type proteins were purified from cultures of *E. coli* HB101 carrying the plasmids, as described previously<sup>30</sup>, except that the cultures were grown at 30 °C rather than 37 °C. Unfolding reaction mixtures (0.1 M Tris-HCl, pH 8.7,



0.2 M KCl, 1 mM EDTA, 30 μM protein and 1 mM DTT<sub>SH</sub>) were kept under N<sub>2</sub>. To quench unfolding samples were reacted with iodoacetate (0.1 M) for 2 min at room temperature and electrophoresed through non-denaturing 15% polyacrylamide gels. Gels were stained with Coomassie blue and relative band intensities determined by scanning with a microdensitometer. The curves in *b* were generated by numerical integration of the rate expressions describing the scheme represented by Equation 1.

were then identified by comparing their electrophoretic mobilities, obtained by non-denaturing gel electrophoresis, with those of wild-type BPTI<sup>32</sup>.

Some of the largest differences among the wild-type and mutant proteins were seen in the kinetics of unfolding (Fig. 2). When the native wild-type protein is incubated under conditions favourable to unfolding (for example, 1 mM DDT<sub>SH</sub><sup>32</sup>) one of the three disulphides of the native protein (14–38) is rapidly reduced. The protein retains, however, most of its native conformation and is an active trypsin inhibitor<sup>33</sup>. Further reduction and unfolding occur by means of intramolecular rearrangements with half-times of about 10 hours<sup>34</sup>. By contrast, the proteins with substitutions at Tyr 23 or Asn 43 unfolded to completion within 10 minutes under these conditions.

The Tyr 35 → Gly mutant behaved differently from the proteins altered at Tyr 23 or Asn 43. Like the wild-type protein, one disulphide of the native protein was rapidly reduced, after which further unfolding was very slow. Because the rate of the subsequent step in unfolding remained slow for this mutant, it is probable that the protein retained much of its native conformation and that the selectively reduced disulphide was 14–38, as in the wild-type protein. Unlike the wild-type protein, the selectively reduced form of the Tyr 35 → Gly mutant was not an active trypsin inhibitor and was rapidly cleaved by trypsin (Fig. 3), indicating that the substitution increases the flexibility of the native-like intermediate.

The kinetics of refolding of the mutant proteins were studied in the same way, except that the reactions were started with the reduced proteins, and the concentrations of DTT<sub>SH</sub><sup>32</sup> and DTT<sub>S</sub> were adjusted to favour refolding. The largest effect on the kinetics of refolding was caused by the Tyr 35 → Gly replacement. Unlike the wild-type and other mutant proteins, this protein did not completely refold in the presence of 40 mM DTT<sub>S</sub>. Instead, a two-disulphide intermediate (with the same electrophoretic mobility as the inactive intermediate formed in the unfolding of this protein) accumulated. This species did form a third disulphide bond to regenerate the native protein in the presence of a stronger oxidizing agent (0.1 mM oxidized glutathione).

The kinetics of unfolding and refolding were analysed quantitatively by comparing the time-dependent changes in the concentrations of the native, reduced and intermediate forms with those predicted by kinetic models. A native-like two-disulphide species (designated II<sub>N</sub>) appeared as the initial intermediate in the unfolding of all the mutant proteins. Other two-disulphide intermediates with lower electrophoretic mobilities, indicative of less compact conformations, were also detected during some of the folding and unfolding reactions. For each mutant protein,

the relative concentrations of these less compact intermediates remained constant during folding and unfolding and consequently these species were treated as a single kinetic class (designated II).

The II<sub>N</sub> species did not, however, seem to be in rapid equilibrium with the other two-disulphide intermediates and was treated as a distinct kinetic class. Accordingly, the following scheme was used as a basis for analysing the folding and unfolding of all of the mutant proteins:



where R and N are the reduced and native states, respectively, and I is a population of one-disulphide intermediates that are assumed to be rapidly interconverting. This scheme corresponds to the main productive pathway for the folding of wild-type BPTI, excluding the steps involving the native-like species with the thiols of cysteines 30 and 51 buried<sup>22,35</sup> (no off pathway species with buried thiols were detected in the folding of any of the mutant proteins).

It was possible to simulate the kinetics of both folding and unfolding using the scheme of Equation 1 and a single set of rate constants for each mutant protein, (Table 1). Examples of the observed and simulated kinetics are shown in Figs 2b and 4. The largest differences between the observed and predicted concentrations corresponded to differences in rate constants of twofold or less.

Although the scheme of Equation 1 provides an adequate description of the kinetics of folding and unfolding these experimental data do not exclude contributions from other pathways, such as the direct formation of II<sub>N</sub> from the one-disulphide intermediates. This is particularly true for the Tyr 23 → Leu and Asn 43 → Ala mutants, for which the interconversions between II and II<sub>N</sub> were not much slower than the timescale of the experiments. Kinetic simulations indicated, however, that the estimates of the relative stabilities of the intermediates derived from the present experiments would not be affected significantly by a contribution from a direct mechanism.

### Mutational effects on folding energetics

The folding mechanisms of the four mutant proteins seem to be qualitatively similar to that of the main productive folding pathway for the wild-type protein. Although the disulphide bonds present in the intermediates for the mutant proteins have not yet been identified, the rate constants can be used to compare the folding energetics of the different proteins. For this purpose, analogous folding steps are compared, recognizing that the amino-acid substitutions may change the distribution of the

TABLE 1 Rate constants for folding and unfolding of wild-type and mutant forms of BPTI

| Protein      | R ↔ I                                         |                                               | I ↔ II                                        |                                               | II ↔ II <sub>N</sub>          |                               | II <sub>N</sub> ↔ N                           |                                               |
|--------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-------------------------------|-------------------------------|-----------------------------------------------|-----------------------------------------------|
|              | Forward<br>(s <sup>-1</sup> M <sup>-1</sup> ) | Reverse<br>(s <sup>-1</sup> M <sup>-1</sup> ) | Forward<br>(s <sup>-1</sup> M <sup>-1</sup> ) | Reverse<br>(s <sup>-1</sup> M <sup>-1</sup> ) | Forward<br>(s <sup>-1</sup> ) | Reverse<br>(s <sup>-1</sup> ) | Forward<br>(s <sup>-1</sup> M <sup>-1</sup> ) | Reverse<br>(s <sup>-1</sup> M <sup>-1</sup> ) |
| Wild-type*   | 0.022                                         | 20                                            | 0.03                                          | 250                                           | 5.0 × 10 <sup>-3</sup>        | 2.0 × 10 <sup>-5</sup>        | 5.7                                           | 30                                            |
| Tyr 23 → Leu | 0.02                                          | 95                                            | 0.012                                         | 250                                           | 2.0 × 10 <sup>-2</sup>        | 1.0 × 10 <sup>-1</sup>        | 0.8                                           | 25                                            |
| Tyr 35 → Gly | 0.03                                          | 50                                            | 0.03                                          | 200                                           | 1.0 × 10 <sup>-2</sup>        | 1.8 × 10 <sup>-5</sup>        | 0.02†                                         | 1250                                          |
| Asn 43 → Gly | 0.05                                          | 30                                            | 0.07                                          | 300                                           | 1.5 × 10 <sup>-2</sup>        | 1.0 × 10 <sup>-2</sup>        | 4.0                                           | 140                                           |
| Asn 43 → Ala | 0.03                                          | 20                                            | 0.03                                          | 200                                           | 1.0 × 10 <sup>-2</sup>        | 2.0 × 10 <sup>-2</sup>        | 0.5                                           | 200                                           |

The rate constants for the various steps in folding and unfolding in the presence of oxidized and reduced DTT were determined by simulation of the observed kinetics with models based on the scheme of Equation 1. For each protein, the rate constants were adjusted until a single set of rate constants predicted all of the folding and unfolding data.

\* The rate constants for the wild-type protein were determined previously<sup>22</sup>.

† The rate constant for the formation of the last disulphide of the Tyr 35 → Gly mutant was too low to be measured directly with oxidized DTT. Its stability was therefore determined by measuring the equilibrium constant for disulphide formation with oxidized and reduced glutathione. The equilibrium constant for forming the disulphide with dithiothreitol was then estimated by dividing the glutathione equilibrium constant (1.9 × 10<sup>-2</sup> M) by 1,200 M, the equilibrium constant for forming the disulphide of oxidized dithiothreitol with glutathione. The forward rate constant for forming the protein disulphide with DTT was then estimated by multiplying the equilibrium constant (1.6 × 10<sup>-5</sup>) by the measured rate constant for the reduction of the disulphide with DTT.



different disulphide-bonded species in the various kinetic classes. The measured rate constants are estimated to be accurate within a factor of two, corresponding to errors in relative free energies of 0.4 kcal mol<sup>-1</sup> or less.

The energetics of the folding transitions of the wild-type and mutant proteins are compared in Fig. 5a and b. The reaction profiles shown in Fig. 5a were calculated using the same thiol-disulphide redox potential, for which the apparent free energy change for wild-type protein folding is -7 kcal mol<sup>-1</sup>. Each of the profiles in Fig. 5b was calculated for a different redox potential, chosen so that the native and reduced states of each protein have equal apparent free energies. In both sets of reaction profiles, the stabilities of the intermediate and native states of each protein are expressed relative to the reduced state of

that protein. Because the different proteins are not reversibly interconvertible their energies cannot be compared directly. The maxima in the reaction profiles represent the apparent free energies of the transition states for the intramolecular steps in disulphide formation or rearrangement. For all mutants, the slowest intramolecular step in folding is the rearrangement of II to II<sub>N</sub>, as it is for the wild-type protein.

Although all four of the substitutions destabilize the native protein (relative to the reduced state), the mutations at the three sites have different effects on the relative stabilities of the folding intermediates. To compare the effects of the substitutions on the different intermediates, the changes in relative stability were normalized by calculating a 'relative destabilization' for each intermediate, (i):

$$\text{Relative destabilization} = \frac{\Delta G_{\text{mutant}}^{\text{R-i}} - \Delta G_{\text{wild-type}}^{\text{R-i}}}{\Delta G_{\text{mutant}}^{\text{R-N}} - \Delta G_{\text{wild-type}}^{\text{R-N}}} \quad (2)$$

The terms  $\Delta G^{\text{R-i}}$  and  $\Delta G^{\text{R-N}}$  represent the free energy differences between the reduced and intermediate states or the reduced and native states, for the mutant and wild type proteins, as indicated by the subscripts. Equation 2 allows the effects of a substitution on the different steps in the folding pathway to be compared, without assumptions being made about how the substitution affects the absolute free energies of the various species.

The relative destabilizations of the intermediates are plotted in Fig. 5c. Although the Tyr 35 → Gly substitution destabilizes the native state by 5.4 kcal mol<sup>-1</sup>, it does not destabilize significantly any of the intermediates. It seems that Tyr 35 therefore contributes to only the last step of folding. The two substitutions at Asn 43 destabilize the native-like two-disulphide intermediate, II<sub>N</sub>, as well as the native protein, but do not affect the earlier intermediates. In addition to destabilizing N and II<sub>N</sub>, the Tyr 23 → Leu substitution also destabilizes the other class of two-disulphide intermediates, II, (by 1.5 kcal mol<sup>-1</sup>) and the one-disulphide intermediates (by 0.9 kcal mol<sup>-1</sup>). Thus, of the amino-acid replacements examined in this study, the change at Tyr 23 has the earliest effect on stability during folding.

The effects of the substitutions on the energies of the transition states can be analysed in a similar way (Fig. 5d). All the substitutions destabilize the transition state for the final step in folding, the conversion of II<sub>N</sub> to N, indicating that all the altered residues contribute to stability by the time this transition state is reached during folding. In contrast, none of the replacements greatly destabilizes the transition state for the rearrangement of II to II<sub>N</sub>, indicating that interactions at all of the altered sites are much weaker in this transition state than in the native protein. This is consistent with the proposal that the formation of this transition state during unfolding involves considerable distortion of the native structure<sup>6</sup>.

The Asn 43 → Gly substitution seems actually to stabilize the transition states for the first three steps in folding. It is conceivable that this substitution enhances the rates of these steps by increasing the flexibility of the polypeptide. This cannot be a general effect of introducing Gly residues, however, as the Tyr 35 → Gly replacement does not affect the rates of these steps.

The smallest mutational effects were on the rate of forming the one-disulphide intermediates from the reduced protein. In the wild-type protein this step involves nearly random pairing of the six cysteine thiols in the unfolded protein<sup>36</sup>. The observation that the rate is not changed by more than twofold indicates that the amino-acid substitutions do not significantly perturb the distribution of conformations making up the unfolded state.

Because the substitutions tend to destabilize the native state more than the intermediates, they reduce the cooperativity of the folding transition. This is illustrated in Fig. 5b where the reaction profile for each of the proteins is shown for conditions where the native and reduced states have equal energies. For the wild-type protein, the intermediates are less stable than the

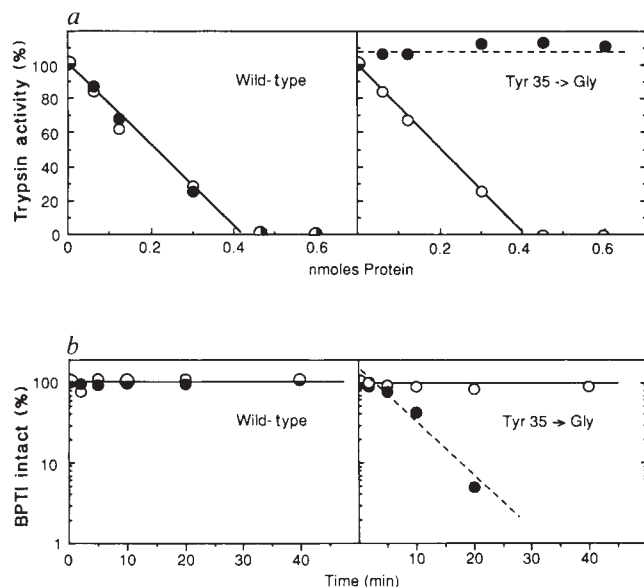


FIG. 3 Trypsin inhibitor activity and trypsin cleavage of the native (N, open circles) and native-like two-disulphide (II<sub>N</sub>, black circles) forms of wild-type and Tyr 35 → Gly BPTI. The two-disulphide forms of the wild-type and mutant proteins were generated by selectively reducing the native proteins with DTT<sub>SH</sub>, as in Fig. 2. *a*, Trypsin inhibition. The indicated amounts of the native or two-disulphide proteins were incubated with 0.5 nmol bovine trypsin for 20 min. The remaining trypsin activity was then determined spectrophotometrically. With all three disulphide bonds intact, both the wild-type and mutant proteins inhibited trypsin stoichiometrically. When the proteins were selectively reduced, the wild-type protein retained its activity, but the mutant was inactive. *b*, Cleavage of the proteins with trypsin. The proteins were mixed with bovine trypsin at final concentrations of 30 μM BPTI and 1.5 μM trypsin. At the indicated times, samples were withdrawn and the trypsin inactivated by adding SDS and heating in a boiling water bath. Samples were electrophoresed through SDS-polyacrylamide gels and the remaining protein concentrations determined by densitometry of the Coomassie blue-stained gels. There was no detectable cleavage of the native wild-type or mutant proteins or the selectively reduced wild-type protein, but the two-disulphide form of the mutant protein was completely degraded in 40 min. METHODS. In *a*, native proteins (concentration; 30 μM) were incubated for 30 min under the conditions described in Fig. 2, with and without 0.25 mM DTT<sub>SH</sub>. Gel electrophoresis of samples reacted with iodoacetate indicated that treatment with DTT<sub>SH</sub> resulted in conversion of 80% of the wild-type protein and 100% of the mutant protein from N to II<sub>N</sub>. The proteins were then diluted and mixed with trypsin to yield the indicated amounts of BPTI and 0.5 nmol trypsin in 0.2 ml of 0.05 M Tris-HCl pH 8.7, 0.1 M KCl, 0.5 mM EDTA and 10 mM CaCl<sub>2</sub>. The mixtures were incubated for 20 min at 25 °C to allow the inhibitor-trypsin complex to form. The remaining free trypsin was then determined from the rate of cleavage of *N*-benzoyl-L-arginine-*p*-nitroanilide at 25 °C, pH 7.8 (ref. 45). In *b*, the native proteins (concentration; 60 μM) were treated for 20 min under the conditions described in Fig. 2, with and without 1 mM DTT<sub>SH</sub>. The samples were then mixed with an equal volume of 6 μM trypsin, 20 mM CaCl<sub>2</sub> and incubated at 25 °C. At the indicated times, samples were withdrawn, mixed with sample buffer for SDS gel electrophoresis (to yield 2% SDS) and heated for 5 min in boiling water. Samples electrophoresed through 10–20% SDS polyacrylamide gels<sup>46</sup>.

end states, and folding corresponds to a two-state transition of the type seen for most small monomeric proteins<sup>37</sup>. For the Tyr 35 → Gly mutant, however,  $II_N$  is 2.4 kcal mol<sup>-1</sup> more stable than both the native and reduced states at the mid-point. Thus, a compact but inactive intermediate (Fig. 3) is the only species that can be easily detected at the mid-point. Similar, but less pronounced effects are seen for the other mutants. These results indicate that cooperativity of folding transitions may not be an inherent property of globular proteins, although it may be correlated with the net stability of the native state<sup>38,39</sup>.

## Conclusions

These experiments show that destabilizing amino-acid replacements at different sites in a protein can have distinct effects on the kinetics and equilibria of different steps in a protein folding reaction. The most striking difference is between the Tyr 35 → Gly mutant and the mutants with changes at Tyr 23 and Asn 43. The Tyr 35 → Gly replacement affects only the last step in folding, whereas the other substitutions largely affect the preceding step,

the conversion of II to  $II_N$ . The proteins with changes at Tyr 23 and Asn 43 can be distinguished further on the basis that the Tyr 23 → Leu substitution has a greater effect on the earlier intermediates (I and II).

Some of the mutational differences can be rationalized by considering the process of unfolding. Tyr 35 is relatively close to the 14–38 disulphide bond in the native structure of the wild-type protein. The removal of the side chain of Tyr 35 favours reduction of the 14–38 disulphide bond and makes  $II_N$  much more susceptible to cleavage by trypsin. It does not, however, significantly affect the rate of further unfolding, which involves intramolecular rearrangements of a disulphide at the other end of the folded protein. In contrast, the replacements at Tyr 23 and Asn 43, which are near the 30–51 and 5–55 disulphides, greatly facilitate the reduction of these disulphides. That the effects of the Tyr 35 → Gly substitution seem to be so well localized in the native conformation indicates that the distances over which different interactions influence one another may be relatively limited.

FIG. 4 Kinetics of refolding of Tyr 23 → Leu (squares, dotted lines), Tyr 35 → Gly (Triangles, short-dashed lines) and Asn 43 → Gly (circles, long dashed lines) BPTI in the presence of 40 mM DTT<sub>S</sub>. The kinetics of refolding of the mutant proteins were measured as in Fig. 2, except that the reduced forms of the proteins were used as the starting materials and DTT<sub>S</sub> was used to promote disulphide formation. Under these conditions, the third disulphide of Tyr 35 → Gly BPTI did not readily form and the protein accumulated as a native-like two-disulphide form ( $II_N$ ). The folding of Tyr 23 → Leu BPTI was also slow because this mutation destabilized intermediates I and II. Both these mutants refolded efficiently in the presence of 0.1 mM oxidized glutathione, a much more potent oxidizing agent.

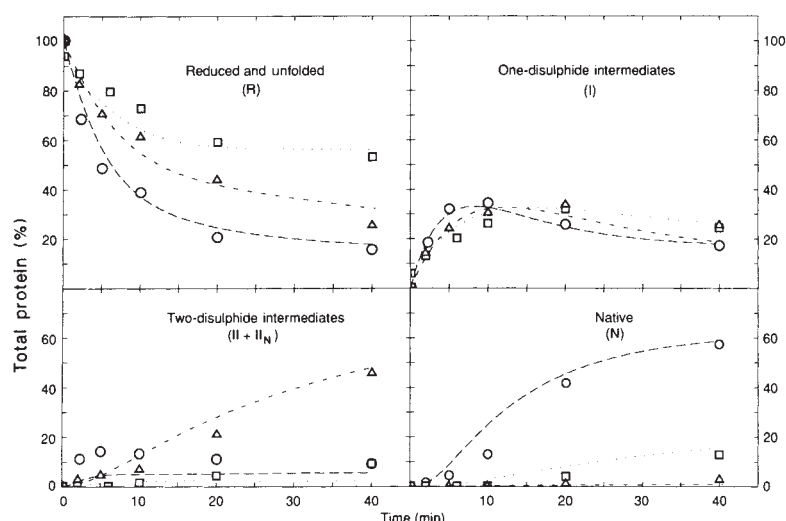
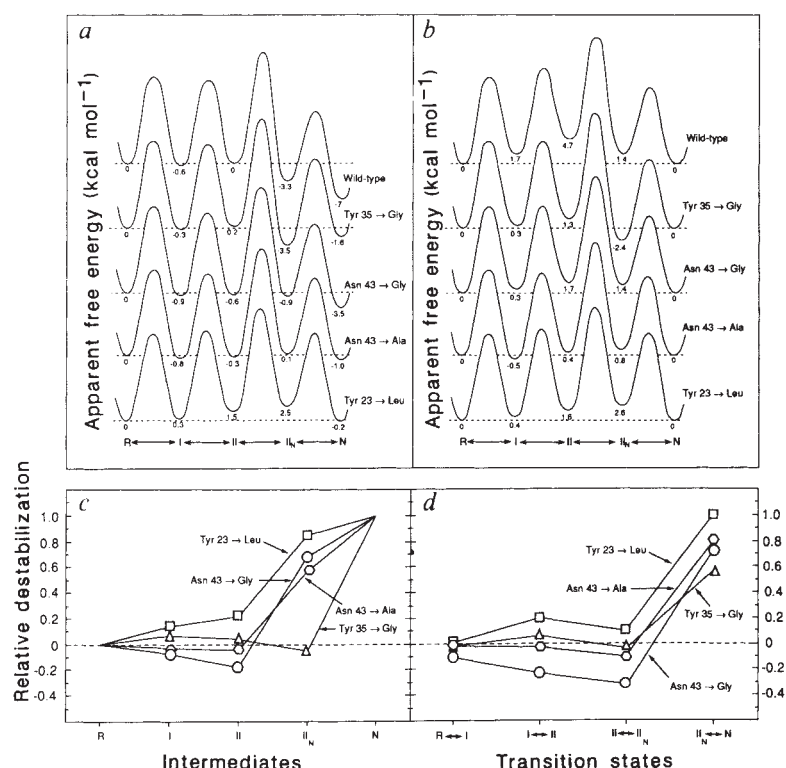


FIG. 5 Effects of the mutations on the energetics of folding. **a**, Reaction profiles for the wild-type and mutant proteins at a single thiol-disulphide redox potential. The relative apparent free energies of the intermediates and transition states in the folding of the different proteins were calculated from the rate constants listed in Table 1, as discussed previously<sup>22,34</sup>. The relative stabilities of the intermediates were calculated for a ratio of DTT<sub>S</sub>:DTT<sub>SH</sub> of 2,700, where the native wild-type protein is 7 kcal mol<sup>-1</sup> more stable than the reduced state. The stabilities of the various species making up the folding transitions of each protein are all expressed relative to the reduced state of that protein. **b**, Reaction profiles for the wild-type and mutant proteins at their mid-point redox potentials, where the native and reduced states of the indicated protein have equal energies (DTT<sub>S</sub>:DTT<sub>SH</sub> = 54 for wild-type, 1,000 for Tyr 35 → Gly, 390 for Asn 43 → Gly, 1,500 for Asn 43 → Ala, and 2,500 for Tyr 23 → Leu). **c**, Relative destabilizations of the folding intermediates. The effects of the indicated amino-acid replacements on the stabilities of the different folding intermediates are compared by normalizing the destabilization of each intermediate (relative to the unfolded state) by the destabilization of the native state caused by the same mutation, according to Equation 2. **d**, Relative destabilizations of the transition states, calculated as in **c**.





The observation that substitutions at different sites influence the various steps in folding differently lends support to folding models in which different regions of the polypeptide acquire structure in a preferred order. The stabilities of the intermediates seem, however, to become more sensitive to each of the amino-acid replacements as the protein approaches the native conformation. This pattern probably reflects the increase in cooperativity among interactions which occurs as more of the native structure forms<sup>40,41</sup>. It seems therefore, that two factors contribute to stability as folding proceeds. First, the number of residues involved in structure increases; and second, the contributions of individual interactions increase as the cooperativity among them increases.

Our results suggest a tentative 'order of action' for three

residues, Tyr 23, Asn 43 and Tyr 35. In the native protein, Tyr 23 forms part of the interface between the  $\alpha$ -helix and  $\beta$ -sheet. This interface also includes the 30-51 disulphide, which is present in the major one-disulphide intermediate and all the productive two-disulphide intermediates in the folding of the wild-type protein. The effects of the Tyr 23  $\rightarrow$  Leu substitution on the stabilities of the early intermediates indicate that the interactions between the helix and sheet may form at an early stage of folding. Recent experiments with peptide fragments of BPTI also indicate that the (30-51) intermediate has considerable native-like structure in this region<sup>42</sup>. Additional studies of these mutant proteins and proteins with substitutions at other sites should further define the roles of different regions of the protein in the folding process.  $\square$

Received 21 November 1988; accepted 17 January 1989.

- Kim, P. S. & Baldwin, R. L. *A. Rev. Biochem.* **51**, 459-489 (1982).
- Creighton, T. E. *J. phys. Chem.* **89**, 2452-2459 (1985).
- Harrison, S. C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028-4030 (1985).
- Scheraga, H. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5740-5744 (1987).
- Creighton, T. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5082-5086 (1988).
- Goldenberg, D. P. & Creighton, T. E. *Biopolymers* **24**, 167-182 (1985).
- Pitts, O. B. & Finkelstein, A. V. *Q. Rev. Biophys.* **13**, 339-386 (1980).
- Karplus, M. & Weaver, D. L. *Nature* **260**, 404-406 (1976).
- Dill, K. A. *Biochemistry* **24**, 1501-1509 (1985).
- Goldenberg, D. P. *A. Rev. biophys. Chem.* **17**, 481-507 (1988).
- Beasty, A. M. *et al. Biochemistry* **25**, 2965-2974 (1986).
- Craig, S. *et al. J. Mol. Biol.* **185**, 681-687 (1985).
- Goldenberg, D. P., Smith, D. H. & King, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7060-7064 (1983).
- Kelley, R. F. & Richards, F. M. *Biochemistry* **26**, 6765-6774 (1987).
- Perry, K. M. *et al. Biochemistry* **26**, 2764-2682 (1987).
- White, T. B., Berget, P. B. & Nall, B. T. *Biochemistry* **26**, 4358-4366 (1987).
- Yu, M.-H. & King, J. *J. biol. Chem.* **263**, 1424-1431 (1988).
- Alber, T. *et al. Nature* **330**, 41-46 (1987).
- Hawkes, R., Grutter, M. & Schellman, J. *J. mol. Biol.* **175**, 195-212 (1984).
- Shortle, D. & Meeker, A. D. *Proteins* **1**, 81-89 (1986).
- Shortle, D., Meeker, A. K. & Friere, E. *Biochemistry* **27**, 4761-4768 (1988).
- Creighton, T. E. & Goldenberg, D. P. *J. molec. Biol.* **179**, 497-526 (1984).
- Creighton, T. E., Kalef, E. & Arnon, R. *J. molec. Biol.* **123**, 129-147 (1978).
- Kosen, P. A., Creighton, T. E. & Blout, E. R. *Biochemistry* **22**, 2433-2440 (1983).

- States, D. J. *et al. J. Molec. Biol.* **195**, 731-739 (1987).
- Wlodawer, A., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **193**, 145-156 (1987).
- Anderson, S. & Kingston, I. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6838-6842 (1983).
- Creighton, T. E. & Charles, I. G. *Cold Spring Harb. Symp. quant. Biol.* **52**, 511-519 (1987).
- Ghrayeb, J. *et al. EMBO J.* **3**, 2437-2442 (1984).
- Goldenberg, D. P. *Biochemistry* **27**, 2481-2489 (1988).
- Marks, C. B. *et al. J. biol. Chem.* **261**, 7115-7118 (1986).
- Creighton, T. E. *J. molec. Biol.* **87**, 579-602 (1974).
- Kress, L. F. & Laskowski, M. S. *J. biol. Chem.* **242**, 4925-4929 (1967).
- Creighton, T. E. *J. molec. Biol.* **113**, 295-312 (1977).
- States, D. J. *et al. J. molec. Biol.* **174**, 411-418 (1984).
- Creighton, T. E. *J. molec. Biol.* **113**, 313-328 (1977).
- Privalov, P. L. *Adv. Prot. Chem.* **33**, 167-241 (1979).
- Robson, B. & Pain, R. H. *J. Biochem.* **155**, 331-344 (1976).
- Hollecker, M. & Creighton, T. E. *J. molec. Biol.* **168**, 409-437 (1983).
- Creighton, T. E. *Biopolymers* **22**, 49-58 (1983).
- Goldenberg, D. P. *J. Cell Biochem.* **29**, 321-335 (1985).
- Oas, T. G. & Kim, P. S. *Nature* **336**, 42-48 (1988).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488-492 (1985).
- Inouye, S. & Inouye, M. *Nucleic Acids Res.* **13**, 3101-3110 (1985).
- Kassell, B. *Meth. Enzym.* **19**, 844-852 (1970).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).

ACKNOWLEDGEMENTS. We thank Dr K. Thomas for synthesizing oligonucleotides and Drs T. Alber, T. E. Creighton and P. S. Kim for helpful discussions and comments on the manuscript. This work was supported by a grant from the U.S. National Science Foundation to D.P.G.

## LETTERS TO NATURE

### Cosmic structure from radiation-blown bubbles

Craig J. Hogan

Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA

**ABSORBING material in an expanding universe filled with sources of radiation is subject to an instability driven by radiation pressure, known in the optically thin limit as the 'mock gravity' instability<sup>1,2</sup>. The same instability in the optically thick limit takes the form of rapidly growing cavities or bubbles in the absorbing material. No matter how small they are when they begin, such bubbles grow to approach a limiting size which depends only on the ratio of photon pressure to absorber inertia. For a non-primordial submillimeter radiation background as strong as that recently detected by the Berkeley-Nagoya group<sup>3</sup>, the bubbles grow to a co-moving radius of  $\sim 20$  Mpc, comparable in scale to the voids seen in the large-scale galaxy distribution.**

Consider the following idealized model. An expanding universe with scale factor  $a \equiv (1+z)^{-1} \propto t^{2/3}$  is filled with uniformly distributed sources of radiation, for example smoothly distributed stars or decaying particles, emitting a luminosity density  $l$ . Gas with mean density  $\rho_g$  absorbs this radiation within a distance negligible compared with any scale under consideration, but is optically thin to its own re-radiated light. The gas is assumed to cool efficiently to a negligible pressure, and has a low enough density that its gravitation is negligible, so that its motion is dominated by radiation pressure alone. Now consider an evacuated bubble in the gas with co-moving radius  $x$ ;

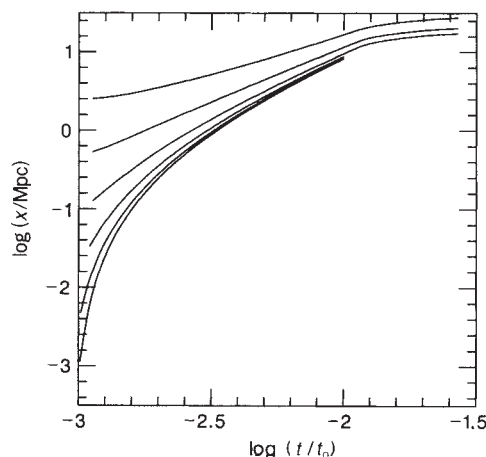
all of the gas inside is swept into a thin wall, leaving the inside filled with radiation sources at the background density. Radiation pressure exerts an outward force on the wall, which moves out, sweeping up more gas as it goes. The wall remains thin because of the efficient cooling, and is assumed to remain spherically symmetrical. The equation of motion for the wall is then<sup>2</sup>

$$\ddot{x} + 2(\dot{a}/a)\dot{x} + 3\dot{x}^2/x - l/\rho_g ca = 0 \quad (1)$$

The bubble wall is accelerated outwards by the radiation pressure of the light from the sources inside, and is slowed both by the cosmic expansion and by the inertia of newly swept-up gas. The mass of the wall increases as  $x^3$ .

To illustrate the general character of the solutions to this equation let us specialize to the case where  $l/\rho_g$  is constant up to a time  $t_1$  and vanishes from then to the present epoch  $t_0$ . For normalization we adopt  $1+z_1=20$  and a value of  $l/\rho_g$  corresponding to a present-day energy density  $\epsilon_0 = 1.0 \times 10^{-13}$  erg cm<sup>-3</sup>, typical of models proposed for the observed submillimetre excess<sup>3,4</sup>, and a density of absorbing material that is 1% of the critical density, typical of estimates of the mean density of galactic material<sup>5</sup>. I assume throughout a Hubble constant of 50 km s<sup>-1</sup> Mpc<sup>-1</sup> and define co-moving lengths by their value at the present epoch. Figure 1 shows numerical solutions of equation (1), starting with zero velocity ( $\dot{x}=0$ ) at  $1+z=100$ , for a variety of initial radii: 0.0008, 0.004, 0.02, 0.1, 0.5 and 2.5 Mpc, spanning a factor of 3,125. The three largest bubbles are shown expanding beyond  $t_1$ , when radiative acceleration stops and the curves level off. After this they continue to sweep up material, but soon approach their final size of  $\sim 20$  Mpc radius, which is slightly larger than the size at  $t_1$ .

FIG. 1 Co-moving bubble radii (in Mpc) as a function of cosmic time (in units of  $t_0$ ) for initial radii at  $(1+z)=100$ , ranging from 0.0008 Mpc to 2.5 Mpc and zero initial velocity. The top three curves extend beyond the switching off of the radiation pressure, where they level off.



It scarcely matters how small the bubbles are initially because they always end up at about the same size. The reason is clear from equation (1): the rate of acceleration relative to the bubble size,  $\ddot{x}/x$ , increases for smaller bubbles, and the braking terms become relatively less important. All solutions converge to the 'attractor' with a response rate of about  $(l/\rho c a x)^{1/2}$ . One can show that in the case of a power-law luminosity  $l/\rho_g \propto (a/a_1)^\gamma$ , the characteristic size of this self-similar attractor at  $t_1$  is<sup>2</sup>

$$x_1 = \frac{3(2\gamma+5)}{(2\gamma+4)(8\gamma+17)} \frac{\rho_r}{\rho_g} c t_0 a_1^{-1/2} \quad (2)$$

where  $\rho_r/\rho_g$  is the final (present-day) ratio of photon density to gas density. The final size is somewhat larger because of the growth after  $t_1$ , although the various solutions no longer converge after this time.

The smallest bubble in Fig. 1 grows by a factor of  $\sim 10^{13}$  in mass between  $z=100$  and the present. Clearly such spectacular amplification in the scale of structure is of interest for explaining the large-scale galaxy distribution, especially as there are (controversial) indications<sup>6</sup> of bubbly or frothy features with characteristic radii of about 25 Mpc in recent redshift surveys. To get such large-scale bubbles, the gas density has to be low. The 1% of the critical value we have adopted is enough to include the luminous parts of galaxies, but corresponds to a mean mass-to-blue-light ratio of only about 10 in solar units. Thus the ubiquitous cosmic dark matter outside galaxies, possibly including remnants of the radiation sources, would not be expected to participate in large-scale clustering, and dark halos of galaxies and clusters would have to form through gravitational accretion onto the clustered gaseous component. (For this to occur it is important that the wall remains gaseous long enough for the deceleration terms to slow it down substantially. In the bubbles shown, the co-moving expansion velocity decreases by a factor of about 10 between  $z=20$  and  $z=10$ ; halo accretion could start by about this time). Note that the radiation instability can contribute to the manufacture of galaxies themselves, even for a cosmic gas density higher than we have assumed by a factor of ten.

Of course the idealizations of the model allow amplifications that could not be achieved in practice; for example, extremely small bubbles would not have optically thick walls, as we have assumed. For a typical atomic continuum opacity in the ultraviolet ( $\sim 10^{-17} \text{ cm}^2$ ), the optical depth of even our smallest bubble is about 20, which is enough for the approximations to hold if the gas is mostly neutral; on the other hand, with this much radiation the neutral fraction soon becomes quite small unless the gas is very clumpy. However, opacities on Mpc scales are expected to be large in a wide variety of situations<sup>2</sup>, making the bubble model superior to linear theory in the context of describing the formation of large-scale structure. Other assumptions of the model, such as the homogeneity of the bubble wall,

the uniform distribution of radiation sources over a time  $t_1$ , and their particular time evolution, certainly overlook important complications, but the goal here (as in other similar models) is a simplified general description. For example, in describing the motion of the bubble wall, it does not much matter whether the luminosity at any particular time is uniformly distributed, as assumed, or is concentrated nearer the edges, as it would be if the age of the evacuated bubble exceeded that of the stars.

Other bubble solutions have been investigated in cosmology, arising as a result of a localized positive energy perturbation, and they also show asymptotic insensitivity to the details of initial conditions<sup>7</sup>. But the final size of those bubbles does depend on the quantity of energy initially deposited. Here the final size depends almost solely on the background density of matter and radiation, which are both global quantities; therefore the final scale is the same regardless of the initial scale of structure. In this feature the present bubbles resemble more closely the macroscopic cavities that form in a first-order phase transition, and whose eventual size depends only on the sound speed of the medium. Of course, if gravity were taken properly into account, the bubbles produced by radiation pressure would eventually (long after the radiation switched off) approach one of the standard self-similar solutions.

Although the bubble solutions have their limitations, they do indicate that the linear theory<sup>1,2</sup> exaggerates the sensitivity of the system both to initial conditions and to the optical depth of the absorbing material. As shown here, nonlinear effects allow the range of the instability to extend far beyond the nominal limit of the linear instability at unit optical depth; they also tend to couple modes very strongly, in the sense that a tiny bubble can directly seed a much larger one. The formal growth rate of these bubbles is also, initially, much greater than the linear growth rate, essentially because a larger opacity per mass is allowed. But the most significant point is perhaps that both the bubble model and the linear theory, two very different approximations, give similar answers for the final scale of structure produced and its dependence on the parameters  $\rho_r/\rho_g$  and  $a_1$ . The essential conclusion remains the same, that if the submillimeter excess is real, then its production almost certainly had a decisive influence on the formation of large-scale cosmic structures.  $\square$

Received 28 November 1988; accepted 19 January 1989.

1. Hogan, C. J. & White, S. D. M. *Nature* **321**, 575-578 (1986).
2. Hogan, C. J. *Astrophys. J.* (in the press).
3. Matsumoto, T. et al. *Astrophys. J.* **329**, 567-571 (1988).
4. Bond, J. R., Carr, B. J. & Hogan, C. J. *Astrophys. J.* (submitted).
5. Cowie, L. L. in *The Post-Recombination Universe* (eds Kaiser, N. & Kasenby, A. N.) 1-18 (Kluwer, Dordrecht, 1988).
6. De Lapparent, V., Geller, M. & Huchra, J. *Astrophys. J.* **302**, L1-L5 (1986).
7. Bertschinger, E. *Astrophys. J. Suppl.* **58**, 1-37 (1985).

ACKNOWLEDGEMENTS. This work was supported by NSF, NASA and an Alfred P. Sloan Foundation Fellowship.



# Disturbed neutral hydrogen in the galaxy NGC3067 pointing to the quasar 3C232

C. L. Carilli\*, J. H. van Gorkom† & John T. Stocke‡

\* National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Department of Astronomy, Columbia University, New York, New York 10027, USA and National Radio Astronomy Observatory

‡ Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, Colorado 80309, USA

THE quasar-galaxy pair 3C232 and NGC3067 was the first in which Ca II absorption was found in the higher redshift spectrum of the quasar at the lower redshift of the galaxy<sup>1</sup>. This has been taken as evidence that quasar absorption lines are due to gas associated with extended disks or spherical haloes of galaxies distributed along the line of sight<sup>2</sup>. Implicit in this is the assumption that redshifts are cosmological. We present new observations of neutral hydrogen at the redshift of NGC3067, both in emission and absorption against the radio-loud quasar. The hydrogen forms a long tail, apparently extending from the galaxy to the quasar and beyond, which seems to be a disturbed extension of one of the galaxy's spiral arms. No companion to NGC3067, which might cause such a disturbance, can be seen. The extended gas distribution, mapped here for the first time in such a system, conforms neither to the extended disk, nor to the spherical halo model. We suggest that most low redshift absorption line systems arise in extended gas associated with interacting galaxies unless, of course, our line-of-sight passes through the optical disk of the associated galaxy.

The optical field for this pair is shown in Fig. 1. The redshift of the quasar is 0.5303, which implies a luminosity distance of 2,760 Mpc (we assume  $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , and  $q_0 = 1/2$ ). The galaxy redshift is 0.0049 (which corresponds to a velocity of  $1,456 \text{ km s}^{-1}$ ), giving a distance of 29 Mpc<sup>1,3</sup>. Notice the prominent dust lanes in the eastern half of the galaxy. The galaxy has been classified as Sc (dust)<sup>4</sup>. Radio spectra of the quasar<sup>3,5</sup> reveal H I 21 cm absorption at a heliocentric velocity of  $1,420 \text{ km s}^{-1}$ , with a full-width half-maximum of  $3.75 \text{ km s}^{-1}$ ; optical spectra (J.T.S., manuscript in preparation) show three heavy-element absorption line systems at  $1,371$ ,  $1,417$  and  $1,545 \text{ km s}^{-1}$ , with widths  $\leq 15 \text{ km s}^{-1}$ . The angular separation of the pair projects to 16 kpc at the redshift of the galaxy. This would be the minimum halo radius if the absorbing gas is halo material. If the absorption were by gas in a flat disk, then the disk radius must be at least 60 kpc, and the rotation velocity at 60 kpc would be twice as high as at the edge of the optical disk of the galaxy<sup>3</sup>.

We made observations on 19 April 1988 for 8 hours in the C configuration of the Very Large Array (ref. 6). Sixty-three spectral channels were used, with a velocity resolution of  $20.8 \text{ km s}^{-1}$ , centred at  $1,456 \text{ km s}^{-1}$  (heliocentric for the 21-cm line) and we used amplitude and phase self-calibration. The quasar continuum was removed according to the guidelines set out in ref. 7. A spectrum of the quasar showed a residual ripple in the band pass with an r.m.s. of 1.6 mJy per beam per channel (about 0.1% of the flux of 3C232). This ripple is large enough to confuse any real emission below about 3 mJy per beam within  $10''$  of the quasar in each spectral channel image.

The systemic velocity for the H I emission from NGC3067 is  $1,465 \pm 10 \text{ km s}^{-1}$ , as calculated from the integrated H I profile, the position velocity profile along the galaxy major axis, and the isovelocity contours. The kinematic centre for the H I is at RA = 09 h 55 min 25.6 s, dec. =  $32^\circ 36' 33''$  (1950; Fig. 2). Hence, the kinematic centre for NGC3067 is displaced from the optical centre by  $10''$  to the west, and from the radio continuum centre

by  $6''$ , again to the west. The galaxy's gaseous disk seems to be truncated in the west just beyond the point where the rotation curve becomes flat. The position-velocity profile along the galaxy major axis (Fig. 3) shows this asymmetry clearly, with the receding side of the disk (east) extending almost twice as far as the approaching side (west).

The most striking feature on our images is the long 'tail' of H I emission extending towards and over the quasar (Fig. 2). The full length of the tail is about  $3'$ , or 24 kpc at the redshift of the NGC3067, and the deconvolved width is about  $17''$ , or 2.3 kpc. The H I distribution in the velocity range from  $1,332$  to  $1,600 \text{ km s}^{-1}$  in the tail is probably gas that has been pulled away from the northern spiral arm of NGC3067. This explains why a rather extended H I disk can be seen south of the galaxy, with no corresponding feature in the north.

Just north of the quasar, the H I emission goes through a minimum and then terminates further to the northeast in a cloud which has a velocity dispersion covering almost the entire velocity range of the galaxy. This emission is suspect because the cloud sits quite closely to the peak sidelobe from 3C232 (about 8% of the peak). If this emission is real, we see an interesting trend for velocity width along the tail. The width of the emission profile increases steadily from  $30 \text{ km s}^{-1}$  in the southwest to  $200 \text{ km s}^{-1}$  in the northeast. Although they are admittedly unreliable, our data at the position of 3C232 fit in quite well with this trend. Other evidence for a large velocity dispersion at

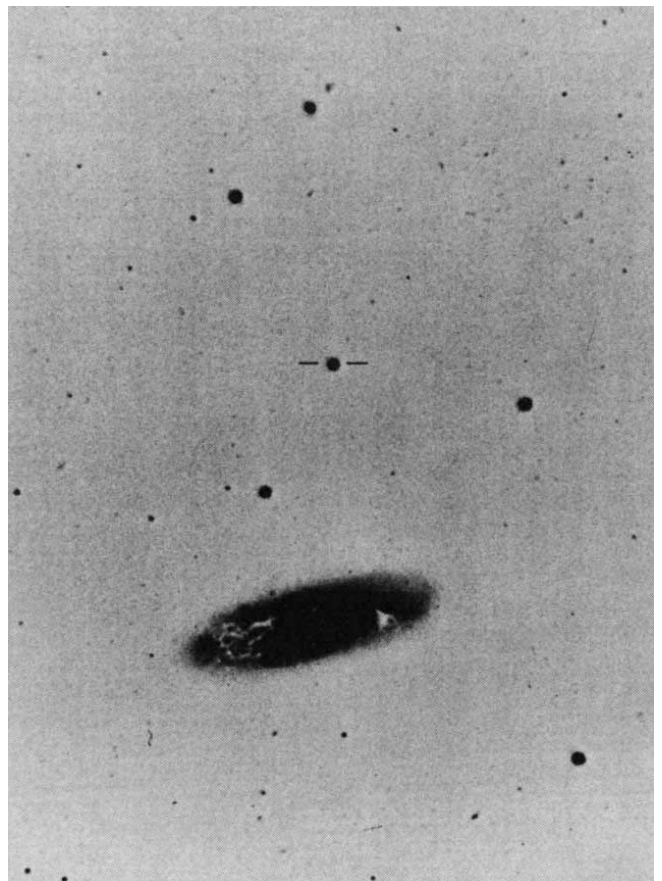


FIG. 1 The optical field of the quasar galaxy pair 3C232-NGC3067 (from a blue plate taken with the Kitt Peak National Observatory 4-m telescope, kindly supplied to us by Dr H. C. Arp). The quasar sits  $2''$  north of the galaxy, and is marked with bars running east-west. The optical radius of the galaxy is  $R_{25} = 74''$ , and the galaxy is inclined at an angle of  $75^\circ$  (ref. 3). The quasar has a continuum radio flux at 1,414 MHz of 1.33 Jy. The integrated continuum flux from NGC3067 is 54 mJy, which implies NGC3067 is probably a starburst galaxy. This assertion is supported by the high  $60 \mu\text{m}$  and  $100 \mu\text{m}$  Infrared Astronomical Satellite fluxes for NGC3067 (9.0 and 18.5 Jy respectively; H. C. Arp, personal communication).

3C232 is the fact that optical absorption occurs over a velocity range of  $170 \text{ km s}^{-1}$ . Note that large velocity dispersions ( $>100 \text{ km s}^{-1}$ ) are often seen in high-redshift ( $z > 0.5$ ) metal-line-absorption systems, while low-redshift systems ( $z < 0.1$ ) typically have small dispersions ( $<50 \text{ km s}^{-1}$ )<sup>8</sup>.

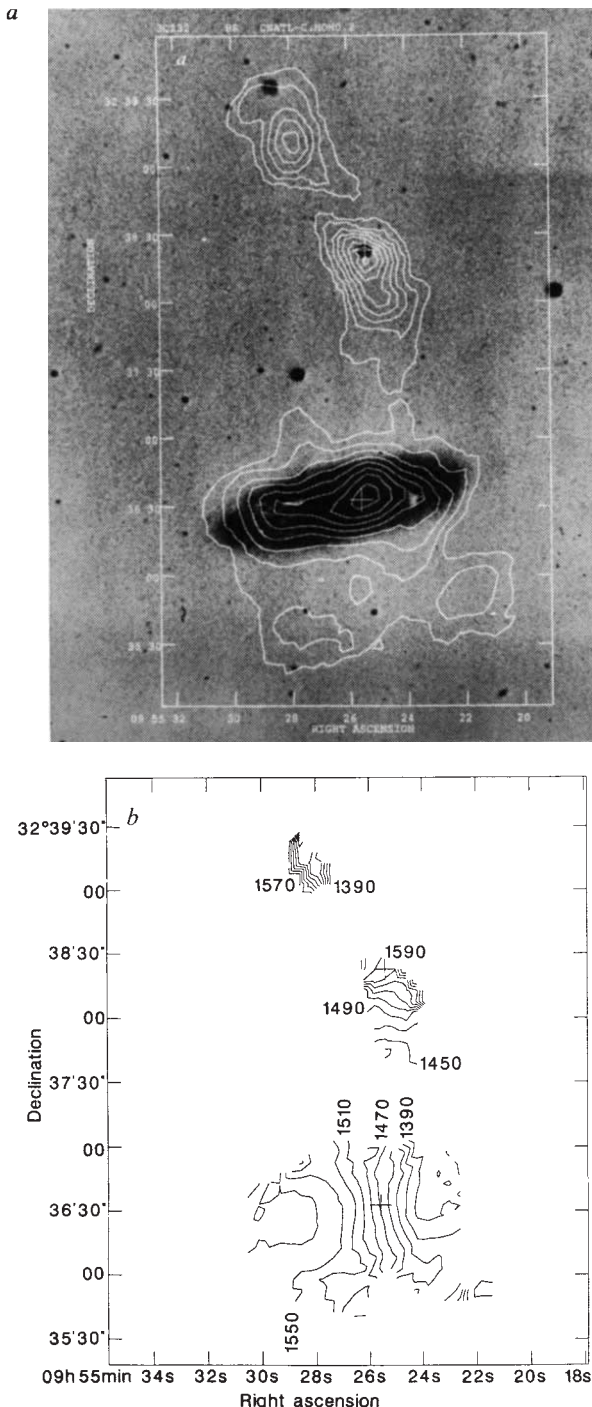


FIG. 2 *a*, The integrated neutral hydrogen emission from NGC3067 (the H I column density and the optical field). Contour levels are: 2.9, 5.9, 10.3, 14.7, 19.2, 23.6, 28.0 and  $32.4 \times 10^{20} \text{ atoms cm}^{-2}$ . *b*, The H I intensity weighted mean velocity field. Contours are spaced by  $20 \text{ km s}^{-1}$ , starting at  $1,370 \text{ km s}^{-1}$ . Crosses mark the quasar and galaxy kinematic centre positions. Notice the systematic deviations from circular motions in the north and south in the isovelocity contours. This is caused by the spiral arms, which are seen in the optical image at these positions. In the east the opening angle of the contours decreases, indicating a flattening of the rotation curve. The peak rotational velocity is  $160 \text{ km s}^{-1}$ . In the west, the position angle of the line of nodes changes dramatically in the outer part, confirming the idea that the galaxy is severely warped in the west.

We assume that the heavy-element absorbing gas is associated with the H I observed in emission. This implies that the absorbing gas associated with NGC3067 towards 3C232 is neither in an extended disk nor a halo, but in a long 'tail' of H I, extending well beyond the galaxy's optical disk. This result is not surprising for a number of reasons. First, observations of 21-cm emission from nearby spirals show that relatively isolated galaxies with regular H I disks rarely show H I extending beyond two Holmberg radii. Recent observations (ref. 9; R. Sancisi, J.H.v.G., T. Cornwell and J. v. Albada, manuscript in preparation) show that for most galaxies, the H I cuts off very sharply near the optical radius of the galaxy, at column densities of  $10^{19} \text{ cm}^{-2}$ . Second, when very extended H I is observed, it is usually in interacting pairs or groups of galaxies (compare the Leo triplet, M51, M348 and most recently Arp 143; refs 10–12). Lastly, from optical absorption studies, Morton *et al.*<sup>13</sup> propose a sharp cutoff for Ca II and Mg II at about the optical radius of a galaxy, although 'the presence of a Magellanic Stream in the halo could increase this distance'. When we review their sources, we find that for every system in their study with known Ca II absorbing gas outside the associated galaxy's optical radius, the galaxy shows evidence that it is involved in an interaction.

Therefore we conclude that the most likely origin for the low-redshift-absorption-line systems (the Ca II systems), is extended gas in interacting systems, unless of course our line-of-sight to a quasar passes through the optical disk of a foreground spiral.

How does our conclusion concerning the low-redshift systems affect our understanding of the origin of the higher redshift quasar-absorption-line systems? Unfortunately, the relationship between the low-redshift systems ( $z_{\text{abs}} \leq 0.1$ ) and the higher redshift systems ( $z_{\text{abs}} \geq 0.5$ ) remains uncertain. On the other hand, systems such as NGC3067 have often been cited as evidence for extended gaseous haloes or disks around galaxies. Our observation shows that this assumption is incorrect. Clearly then, using such systems to support the idea that the high redshift absorption line systems arise in galaxy haloes is also incorrect.

It is clear that NGC3067 has been severely disturbed. This is evident from its optical appearance, its very asymmetric gas distribution, and the presence of the H I tail. We now turn to the question of what object caused this disturbance. The problem is that NGC3067 has no obvious optical companions. An inspection of the Palomar Sky Survey shows a compact galaxy MRK412 about  $10'$  south of NGC3067. But this galaxy has a

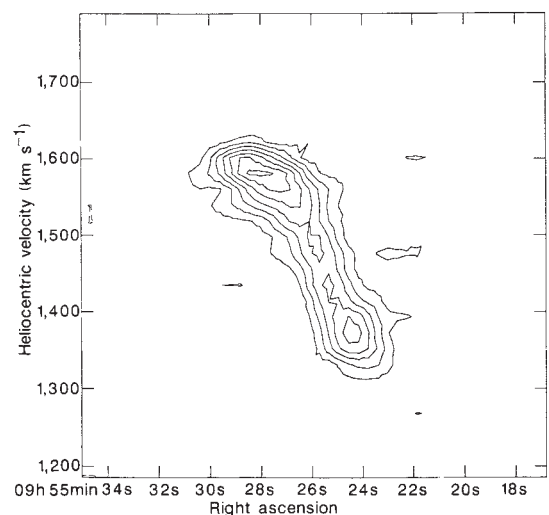


FIG. 3 A position-velocity plot along the major axis of NGC3067 (p.a.,  $101.5^\circ$ ). Notice how the rotation curve flattens in the east at  $\sim 1,600 \text{ km s}^{-1}$  and extends about  $70''$  from the galaxy kinematic centre. In the west the emission terminates just where the flat rotation curve should begin. The truncated disk in the west is also obvious in optical slit spectra along the galaxy major axis<sup>3</sup>.



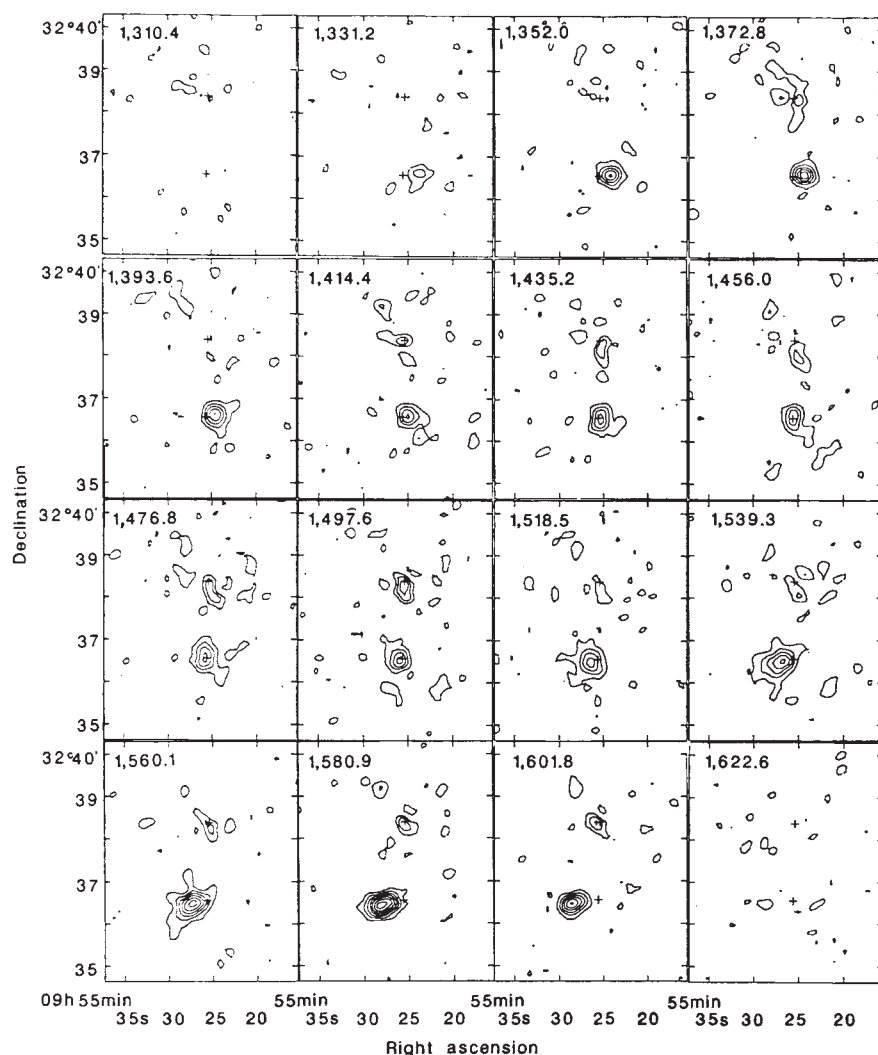


FIG. 4 The central 16 spectral line channels. This covers the entire velocity range of the galaxy ( $1,300$  to  $1,620 \text{ km s}^{-1}$ ). Crosses mark the position of the quasar and galaxy centre. The noise per channel on the natural weighted images was  $0.6 \text{ mJy per beam}$ , or  $1.1 \text{ K}$ . The heliocentric channel velocity for the  $\text{H I}$  is indicated above each image. The contour levels are:  $-4.5, -3.0, -1.5, 1.5, 3.0, 4.5, 6.0, 7.5, 9.0, 10.5, 12.0, 13.5 \text{ mJy per beam}$ . The full-width half-maximum of the gaussian restoring beam after CLEANING was  $19''$  for the natural weighted images. Notice how the tail first appears close to the quasar at  $1,372 \text{ km s}^{-1}$ . Then at  $1,414 \text{ km s}^{-1}$  it is seen in absorption against 3C232, and in emission to the northeast and southwest of the quasar. At  $1,432$  and  $1,456 \text{ km s}^{-1}$  it comes up more strongly close to the galaxy. Images at lower resolution show the tail connecting to the galaxy in this velocity range. With increasing velocity, the tail extends to the northeast while the galaxy's emission moves to the southeast. It can be seen close to 3C232 at velocities as high as  $1,600 \text{ km s}^{-1}$ . Notice how well the tail fits in with the velocity field of the galaxy. In fact, inclusion of the tail in the integrated  $\text{H I}$  spectrum of NGC3067 removes the pronounced asymmetry in velocity that was first noticed by Rubin *et al.*<sup>3</sup>. Both these facts support the conclusion that the tail is  $\text{H I}$  gas associated with NGC3067.

redshift three times larger than NGC3067 (ref. 14). MRK411 has a redshift similar to NGC3067, but it sits  $75'$  north, which projects to about  $0.6 \text{ Mpc}$  at the galaxy redshift. There is also a small spiral galaxy  $15'$  northeast of NGC3067, although its distance is unknown. It is most likely that the disturbing mass has been cannibalized by NGC3067. This conclusion is supported by the appearance of a multiple optical nucleus in the images presented by Burbidge *et al.*<sup>15</sup>. Also, the unusually strong radio continuum and infrared luminosities of NGC3067 place it in the starburst class of galaxies, which are currently thought to be mergers.

We conclude with a brief discussion of the implications this system has on our understanding of quasar redshifts. The position of the quasar in this tail may be a coincidence of projection. We have the additional coincidence that the velocity dispersion at the quasar seems to be considerably larger than that in the tail to the south of the quasar. This is evidenced both by the radio  $\text{H I}$  emission studies and by the optical absorption studies (although we have mentioned potential problems with the radio data at the quasar position). Radio observations with improved band-pass calibration would be extremely useful in clarifying the velocity field in the vicinity of the quasar.

Although reminiscent of the optical examples of tails pointing to quasars<sup>16</sup>, an important difference is that this tail is hydrogen-rich, which allows us to calculate the effects the quasar would have on the ambient material if it were local to the tail. From our  $21\text{-cm}$  observations, and from extrapolation to shorter wavelengths of the International Ultraviolet Explorer spectra

of 3C232, we calculate a radius of  $1.8 \text{ kpc}$  for the  $\text{H II}$  region induced by the quasar. This is slightly larger than the tail radius. The expected emission measure from such a region is  $190 \text{ pc cm}^{-6}$ , which is just below the sky survey limit. One can easily go deeper than the sky survey with even short exposures on metre class telescopes. If no emission is observed, then stringent limits can be set on the quasar-tail distance.  $\square$

Received 28 November 1988; accepted 18 January 1989.

1. Boksenberg, A. & Sargent, W. L. W. *Astrophys. J.* **220**, 42–46 (1978).
2. Weymann, R. J., Carswell, R. F. & Smith, M. G. A. *Rev. Astr. Astrophys.* **19**, 41–76 (1981).
3. Rubin, V. C., Thonnard, N. & Ford, W. K. *Astrophys. J.* **87**, 477–485 (1982).
4. Sandage, A. & Tammann, G. A. *A Revised Shapley-Ames Catalogue of Bright Galaxies* 30–31 (Carnegie Institute, Washington DC, 1987).
5. Haschick, A. D. & Burke, B. F. *Astrophys. J.* **200**, 137–140 (1975).
6. Napier, P. J., Thompson, A. R. & Ekers, R. D. *Proceedings of the Institute for Electronics and Electrical Engineers* **71**, 1295–1320 (1983).
7. van Gorkom, J. H. & Ekers, R. D. in *NRAO Summer School on Synthesis Imaging* (eds Perley, R., Schwab, F. & Bridle, A.) 177–187 (NRAO, Socorro, New Mexico, 1985).
8. York, D. G., Dopita, M., Green, R. & Bechtold, J. *Astrophys. J.* **311**, 610 (1986).
9. Sancisi, R. in *IAU Symposium 100: Internal Kinematics and Dynamics of Galaxies* (ed. Athanassoula, E.) 55–62 (Reidel, Dordrecht, 1983).
10. Haynes, M. P., Giovanelli, R. & Chincarini, G. A. *Rev. Astr. Astrophys.* **22**, 445–470 (1984).
11. Simkin, S. M., van Gorkom, J. H., Hibbard, J. & Su, H.-J. *Science* **235**, 1367–1370 (1987).
12. Appleton, P. N., Ghigo, F. D., van Gorkom, J. H., Schombert, J. & Struck-Marcell, C. *Nature* **330**, 140–142 (1987).
13. Morton, D. C., York, D. G. & Jenkins, E. B. *Astrophys. J.* **302**, 272–279 (1986).
14. Mazzarella, J. M. & Balzano, V. A. *Astrophys. J. Suppl. Ser.* **62**, 751–819 (1986).
15. Burbidge, E. M., Burbidge, G. R., Solomon, P. M. & Strittmatter, P. A. *Astrophys. J.* **170**, 233–240 (1971).
16. Arp, H. C. *Quasars, Redshifts and Controversies* (Interstellar Media, Berkeley, California, 1986).

ACKNOWLEDGEMENTS. The National Radio Astronomy Observatory (NRAO) is operated by Associated Universities Inc., under contract with the National Science Foundation.

# Contrast mechanism for resolving organic molecules with tunnelling microscopy

J. K. Spong\*, H. A. Mizes†, L. J. LaComb Jr‡, M. M. Dovek‡, J. E. Frommer\* & J. S. Foster\*

\* IBM Research Division, Almaden Research Center, San Jose, California 95120, USA

† Xerox Research Center, Webster, New York 14580, USA

‡ Department of Applied Physics, Stanford University, Stanford, California 94305, USA

THE scanning tunnelling microscope (STM) is now an established tool for the study of ordered metal and semiconductor surfaces on an atomic scale<sup>1,2</sup>. Here we demonstrate the ability of the STM to image a large class of organic molecular adsorbates with almost atomic resolution<sup>3</sup>. From the images it is apparent that not only is the STM resolving the individual molecules, but it is also distinguishing between the different functional groups within the molecules. We propose that the contrast mechanism responsible for the well resolved images is modulation of the local work function of the substrate by the polarizable molecular adsorbates.

The STM operates by bringing a sharp electrode ('tip') into close proximity to a flat conducting sample. If the tip-to-sample separation is sufficiently small ( $<10$  Å), electrons can tunnel across to the opposite electrode. The magnitude of the tunnelling current depends sensitively on the work function of the electrodes, and on the distance between them. The tunnelling current can therefore reveal changes of material or topography with very high spatial resolution. Images of a surface can be made by scanning the tip over the sample and monitoring the amount by which the tip must be retracted in order to keep the tunnelling current constant.

The molecules that we have studied are liquid crystals, characterized by their ability to self-assemble into one-dimensionally (nematic) or two-dimensionally ordered (smectic) liquid phases. Samples are prepared for the STM by applying several milligrams of the bulk solid to the surface of freshly cleaved, highly orientated pyrolytic graphite mounted on a heated microscope stage. As the sample melts, molecules from the liquid become physisorbed to the graphite surface in large, well ordered molecular arrays. The tunnelling tip is immersed in the droplet, and brought to within tunnelling range of the adsorbate-covered surface. Imaging is done in the constant current mode, with  $+0.8$  V bias on the tip, and drawing  $0.2$  nA of tunnelling current.

Figure 1 is an STM image of the liquid crystal 5-nonyl-2-nonyloxyphenylpyrimidine (commercially known as PYP 909) adsorbed on the surface of graphite. This molecule has an aromatic core (phenylpyrimidine) disubstituted with an alkyl and an alkoxy side chain. In the STM image, the molecules appear bright on a dark background, and are apparently lying down on the surface and arranged in rows. The bright oval areas correspond to the phenylpyrimidine cores from which the dimer, aliphatic side chains extend. The total corrugation in the image is about  $1$  Å. The row-width of the molecular array is  $25$  Å, which is  $\sim 7$  Å shorter than the molecular length because of a slight overlap of the alkyl side chains. Our interpretation of the molecular-packing arrangement is shown superimposed on the image. The liquid crystal 4-cyano-4'-*n*-butoxybiphenyl (commercially known as M12 or 4OCB) consists of a biphenyl core with a cyano 'head' group and an alkyl or alkoxy 'tail' group. Anomalous periodicities in X-ray diffraction data of cyanobiphenyl compounds have led to the proposal that many cyanobiphenyls form a bilayer in their liquid-crystalline phases<sup>4,5</sup>. The bilayer structure of the liquid crystal 4OCB is confirmed in the STM image of Fig. 2. Rows of molecules,  $25$  Å in width, are apparent in the figure, but within each row

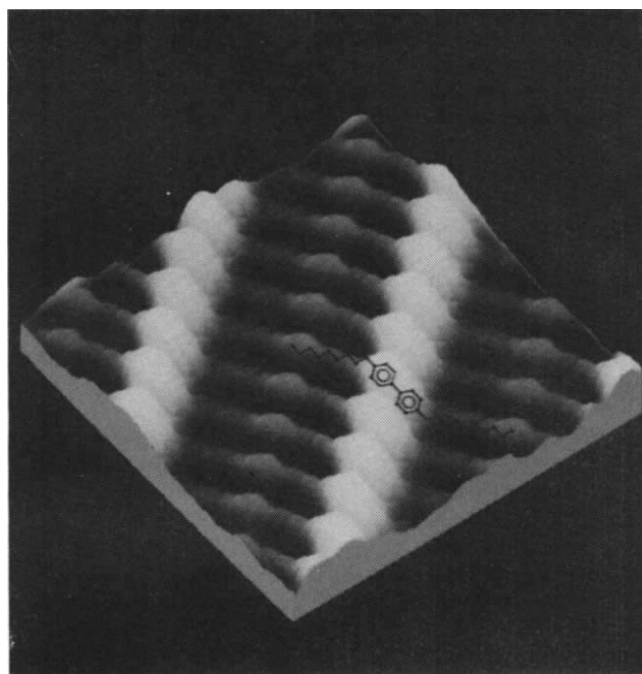


FIG. 1 STM image of PYP 909 on graphite. The scan dimensions are  $43 \times 52$  Å. The three-dimensional view is obtained by assigning to each data point a vertical height that is proportional to its brightness. The data have been unit-cell averaged<sup>17</sup>.

molecules can be seen lying head-to-head against, and interdigitated with, a set of opposing molecules. The biphenyl cores appear brighter than the aliphatic tail groups, and the total corrugation is about  $1$  Å.

The last example, 4-(4'-*n*-propylcyclohexyl)-cyanocyclohexane (CCH3), differs from the previous two in that it contains no oxygen or aromatic structures. It is composed instead of a bicyclohexane core with a cyano head group and a short  $C_3H_7$  aliphatic tail. The STM image of this molecule on graphite (Fig. 3) shows that it also forms a bilayer, and the row width is about  $26$  Å. Although this image is similar to the previous two, we include it to demonstrate that no particular moiety is necessary to image these molecules; the only factor common to all three is that they are all immobilized in a liquid-crystalline matrix.

The STM images of Figs 1–3 reflect the magnitude of the tunnelling current as a function of position, as the tip is scanned over the adsorbed liquid-crystal molecules. These molecules are non-conducting, and are not chemically bound to the surface, so that the contrast mechanism that allows them to be imaged is not obvious. It is doubtful that resonant tunnelling or other forms of electronic spectroscopy are operative, because the occupied ground-state molecular energy levels (and unoccupied levels) lie well below (well above) the substrate Fermi level, and are inaccessible with the small tunnelling voltages we have used here.

Our explanation for the contrast observed in these images focuses on the barrier height, or effective work function of the substrate, as modified by the molecular adsorbate. For small applied bias voltages,  $V$ , the magnitude of the tunnelling current,  $I$ , depends on the potential barrier  $\Phi(z)$  (in eV), according to<sup>6</sup>

$$I \sim V \exp \left( - \int_0^s \{ \Phi(z) \}^{1/2} dz \right) \quad (1)$$

where  $s$  is the tip-to-sample separation (in Å) along the  $z$ -axis. If the integration is carried out to infinity (the 'far-field'), the barrier height becomes the work function of the surface. But at



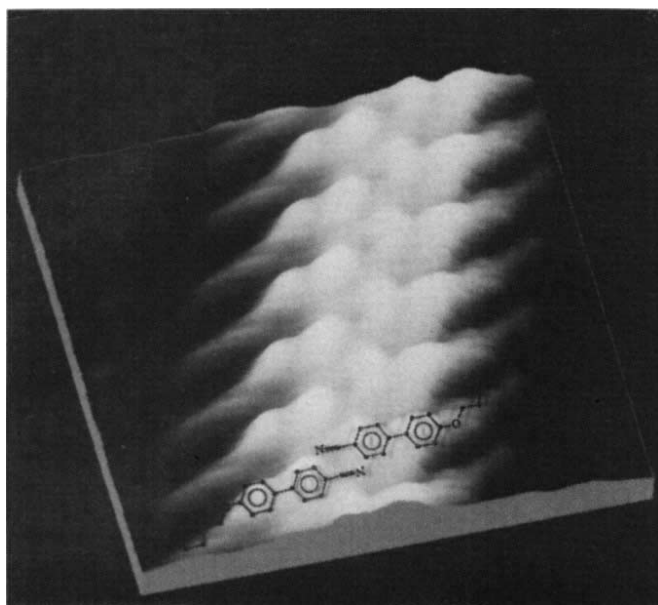


FIG. 2 STM image of 4OCB on graphite. The scan dimensions are  $30 \times 30 \text{ \AA}$ .

the small tip-to-sample separations used in tunnelling microscopy (the 'near-field'), the barrier height  $\Phi(z)$  is substantially lower than the full work function<sup>6</sup>. If  $\Phi(z)$  is a slowly varying function of  $z$ , then the tunnelling current is approximately

$$I \sim V \exp(-\Phi^{1/2}s) \quad (2)$$

where  $\Phi$  is the barrier height at the separation  $s$ . It is well known that the barrier height of a clean surface,  $\phi$ , is altered by the presence of a monolayer of a polar adsorbate according to<sup>7</sup>

$$\Phi = \phi - e\mu/\epsilon_0 \quad (3)$$

where  $\mu$  is the dipole moment density of the adsorbate,  $e$  is the electronic charge and  $\epsilon_0$  is the permittivity of free space. The dipole moment of the adsorbate may be permanent,  $\mu_0$ , or

induced by an electric field  $E$  through its polarizability,  $\alpha$ :

$$\mu = \mu_0 - f(\alpha, E) \quad (4)$$

The function  $f(\alpha, E)$  may include nonlinear effects if the electric field is sufficiently strong. Contributing to the electric field are the image dipole field<sup>8,9</sup>, the depolarizing field of surrounding dipoles<sup>10</sup>, and the tip-to-sample field. Clearly if the polarizability of the adsorbed layer varies as a function of position, then the tunnelling current will also be a function of position. An STM image of such a surface will be a map of changes in the local work function.

We can estimate the change in the barrier height, and therefore the change in the tip-to-sample separation, resulting from the physisorption of benzene on graphite. The work function of clean graphite has been measured as 5.0 eV (ref. 11). We assume that benzene is adsorbed on the graphite surface at an equilibrium distance of 3.3 Å and is bound by 0.5 eV (ref. 12). The interaction of the molecule with its image causes it to acquire a dipole moment of  $\sim 2.2 \times 10^{-29} \text{ C m}$ . The presence of the dipole lowers the work function of an electron at the surface by  $\sim 2.5 \text{ eV}$  (this value is obtained from a correction to equation (3) to account for the finite size of the benzene ring). Experimental values are available for the adsorption of benzene on platinum, and the measured work function changes from 5.4 eV to 3.6 eV on absorption of benzene<sup>13</sup>.

The change in the near-field barrier height can be measured with the STM. By fitting experimental data to equation (2) we find that the barrier height varies periodically from 0.06 eV to 0.1 eV as the tip is scanned over the rows of molecules (the low values are typical of many work-function measurements made by the STM, and are at least in part due to the near-field effects detailed in refs 6, 14 and 15). Therefore, both near-field and far-field estimates of the barrier height indicate that it changes by a factor of about 0.6 as the tip scans over an adsorbed benzene ring.

It is clear from equation (2) that to image a surface in the constant-current mode requires that  $\Phi^{1/2}s = \text{constant}$ . If we assume that the tip is initially 5 Å from the clean graphite surface, it must pull back an additional 1.4 Å when tunnelling over a benzene ring. This number is in reasonable agreement with the 1-Å corrugation observed in the STM images. The alkyl tails of the liquid-crystal molecules are less polarizable and

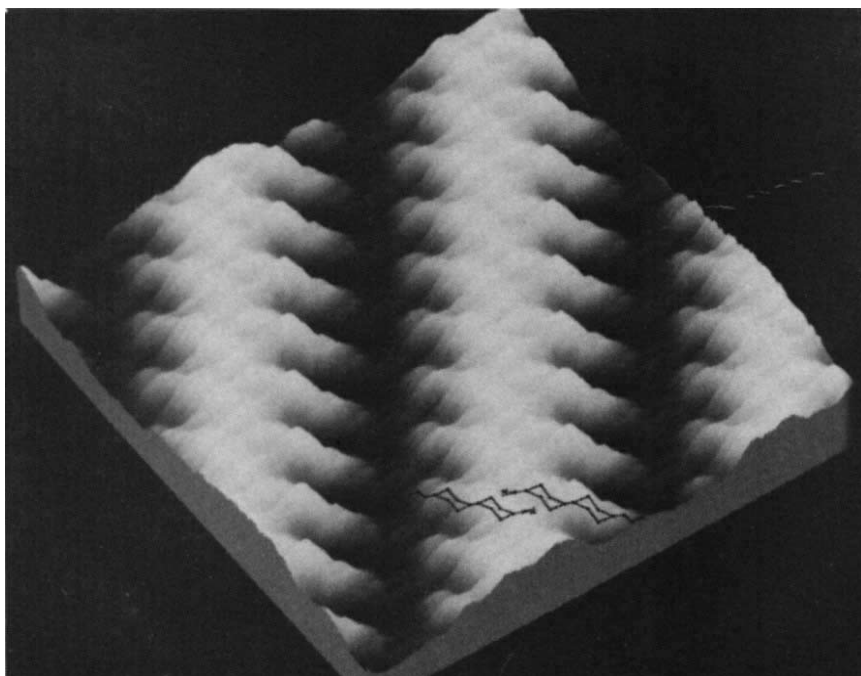


FIG. 3 STM image of CCH3 on graphite. The scan dimensions are  $52 \times 52 \text{ \AA}$ .

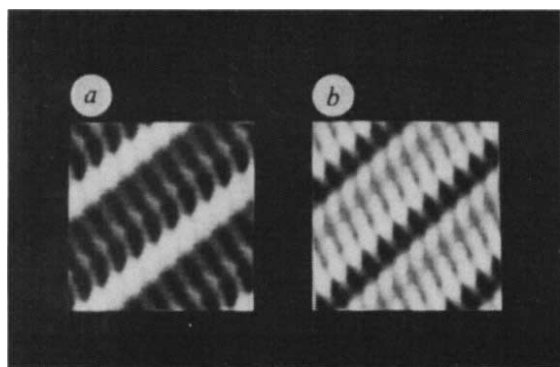


FIG. 4 *a*, STM image of PYP 909, showing the bright aromatic portions of the molecules and the dimmer aliphatic parts. *b*, Identical image of PYP 909, except that the contrast has been reversed. The two images were taken under identical tunnelling parameters, and about 10 min apart.

modulate the work function more weakly. They therefore appear dimmer in the images. Functional groups which are electronically as dissimilar as benzene and cyclohexane can have similar contrast when imaged by the STM, because their polarizabilities are comparable<sup>16</sup>. Finally, the contrast also depends on the size and extent of the dipole layer of the particular substrate and on the ability of the adsorbed molecules to modify it.

We note that according to equations (3) and (4), the observed contrast can change, or even reverse, depending on the sign and magnitude of the external electric field applied by STM tip (we estimate that the field that is due to the STM is of the same order as that from the polarized adsorbate,  $\sim 10^7$  V m<sup>-1</sup>). The STM images generated can therefore depend on tip voltage and mistakenly appear to contain spectroscopic information. The contrast can also change if the tip changes its shape or absorbs dielectric material. This effect is demonstrated in Fig. 4*a* and *b*, which shows STM images of the liquid crystal PYP 909. Figure 4*a* displays the usual contrast seen when imaging this molecule: the central aromatic core appears brighter than the aliphatic tails. For approximately 5% of the time, however, the molecule is imaged with the reverse contrast, as shown in Fig. 4*b*. We attribute the reversal to the temporary adsorption of a molecule on the tip, causing a change in the electric field across the liquid-crystal monolayer.

To date, we have imaged five widely divergent types of liquid-crystal molecules, all with near-atomic resolution (see also ref. 3). We conclude that the contrast mechanism for imaging physisorbed molecules is the modulation of the substrate work function by the polarizable adsorbate. The modulation is sufficiently local to enable the STM to resolve individual molecules spaced by 5 Å and to distinguish between the different functional groups within the molecules. □

Received 25 November 1988; accepted 25 January 1989.

1. Binnig, G. & Rohrer, H. *Helv. phys. Acta* **55**, 726–735 (1982).
2. Binnig, G. & Rohrer, H. *IBM J. Res. Dev.* **30**, 355–369 (1986).
3. Foster, J. S. & Frommer, J. E. *Nature* **333**, 542–545 (1988).
4. Brownsey, G. J. & Leadbetter, A. J. *Phys. Rev. Lett.* **44**, 1608–1611 (1980).
5. Leadbetter, A. J., Frost, J. C., Gaughan, J. P., Gray, G. W. & Mosley, A. J. *Physique* **40**, 375–380 (1979).
6. Lang, N. D. *Phys. Rev. B* **37**, 10395–10398 (1988).
7. Weber, R. E. & Peria, W. T. *Surf. Sci.* **14**, 13–38 (1969).
8. Antoniewicz, P. R. *Phys. Rev. Lett.* **32**, 1424–1425 (1974).
9. Antoniewicz, P. R. *Surf. Sci.* **52**, 703–708 (1975).
10. Topping, J. *Proc. R. Soc. A* **114**, 67–72 (1927).
11. Robrieux, B., Faure, R. & Dussaulcy, J. P. *C.R. Acad. Sci. Ser. B* **278**, 659–662 (1974).
12. Battezzati, L., Pisani, C. & Ricca, F. *J. chem. Soc.* **71**, 1629–1639 (1975).
13. Gland, J. L. & Somorjai, G. A. *Surf. Sci.* **38**, 157–186 (1973).
14. Gimzewski, J. K. & Moller, R. *Phys. Rev. B* **36**, 1284–1287.
15. Lang, N. D. *Phys. Rev. B* **36**, 8173–8176 (1987).
16. *CRC Handbook of Chemistry and Physics* 66th Edn, E-70 (1985).
17. Crepeau, R. H. & Fram, E. K. *Ultramicroscopy* **16**, 7–17 (1981).

ACKNOWLEDGEMENTS. We thank F. Allen, P. Bagus, E. Barrall, P. Hansma, M. Miller, H. Ohtani and C. F. Quate for helpful discussions.

## Evidence for sinking of small particles into substrates and implications for heterogeneous catalysis

P. M. Ajayan & L. D. Marks

Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA

THE shapes and structures of small metal particles (1–10 nm in size) supported on oxide substrates and the effect of substrate-particle interactions in such systems are of substantial importance for heterogeneous catalysis. Here we present, for the first time, experimental evidence using atomic-resolution electron microscopy of gold clusters on MgO smoke particles, which shows that a small particle sitting on a substrate may be thermodynamically unstable and can 'sink' into the substrate. A simple theoretical model is proposed to explain this effect. The results may be of importance in small-particle catalyst systems. The stability of the particle with respect to sinking, and hence reactivity, can be changed by chemisorption of impurities, which will change the surface free energies. A plausible model of catalyst deactivation is that the particle sinks diffusively into the substrate in the course of time. Regeneration of the catalyst could require changing the surface free energy by chemisorption, causing the particle to float back onto the substrate surface.

The structure of a small particle may not resemble the structure of the same bulk material because the large surface-to-volume ratio gives rise to a large surface-energy contribution to the total energy. The shape of a small particle is highly surface-energy-dependent and can be estimated by the Curie-Wulff construction<sup>1,2</sup>. Energetically stable configurations at small sizes take the form of icosahedra and particles that are multiply twinned; as demonstrated recently, the particles can undergo quasi-melting transformations between various forms<sup>3</sup>. What happens when such a particle is placed on top of a relatively large substrate, as one finds, for example, in heterogeneous catalysts? The resulting configurations will be determined by the relative surface free energies of the facets of the particle and the substrate, and the interfacial free energy. Theories for particle shapes on a substrate and the associated wetting transitions have been

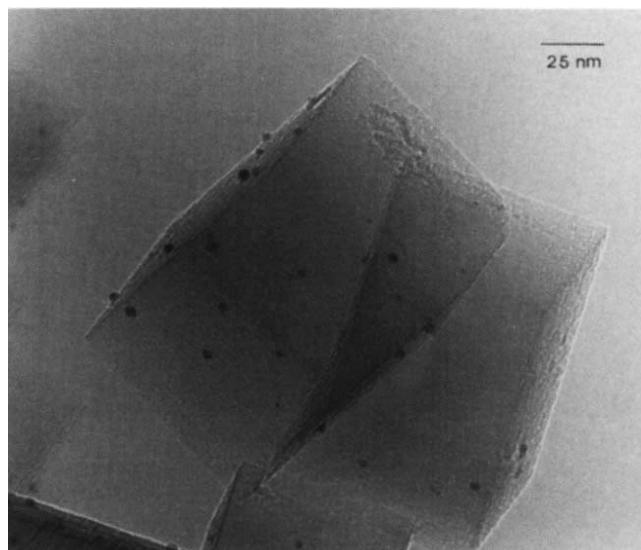


FIG. 1 Low-magnification electron-microscope image showing cuboids of transparent MgO smoke particles with gold microclusters on the surfaces.



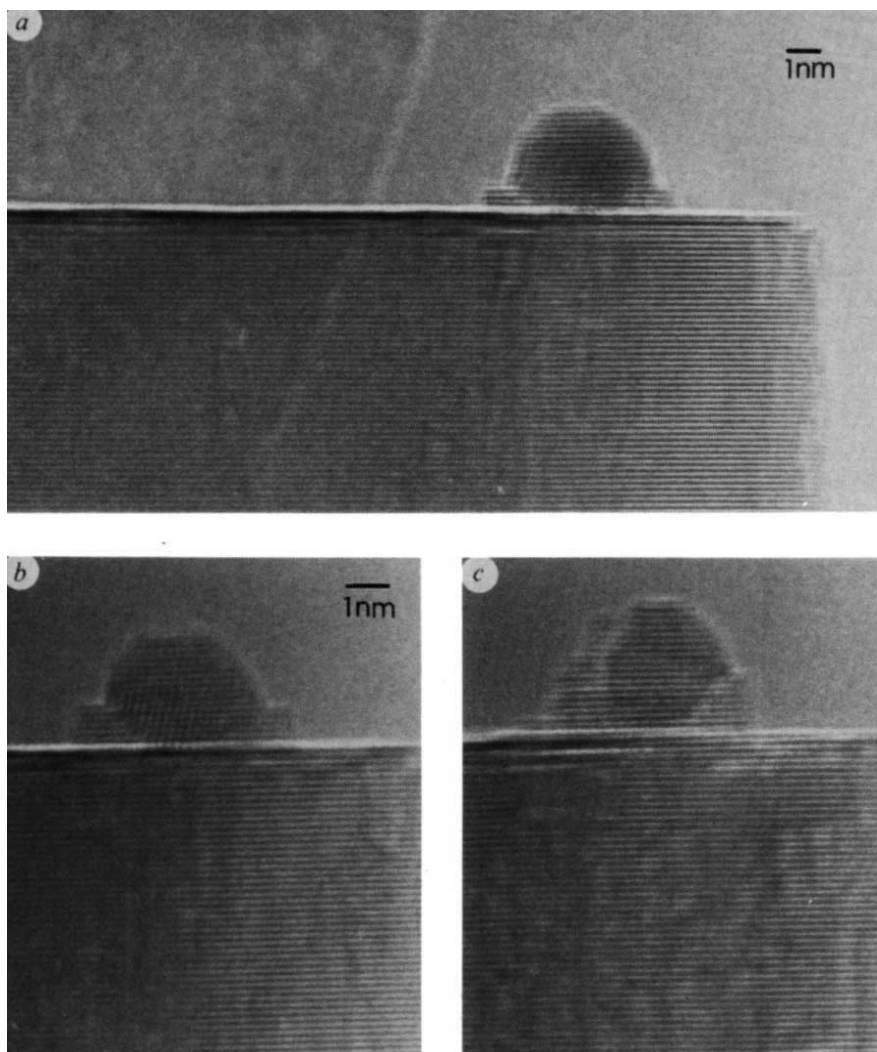


FIG. 2 *a*, High-resolution image of a gold particle on an MgO surface. Layers of MgO form on the particle following irradiation with an electron beam. The lattice layers in the substrate correspond to the (200) atomic planes. *b* and *c* show the same particle after low-flux electron irradiation for 10 and 60 min respectively. The particle sinks into the substrate and also undergoes some shear.

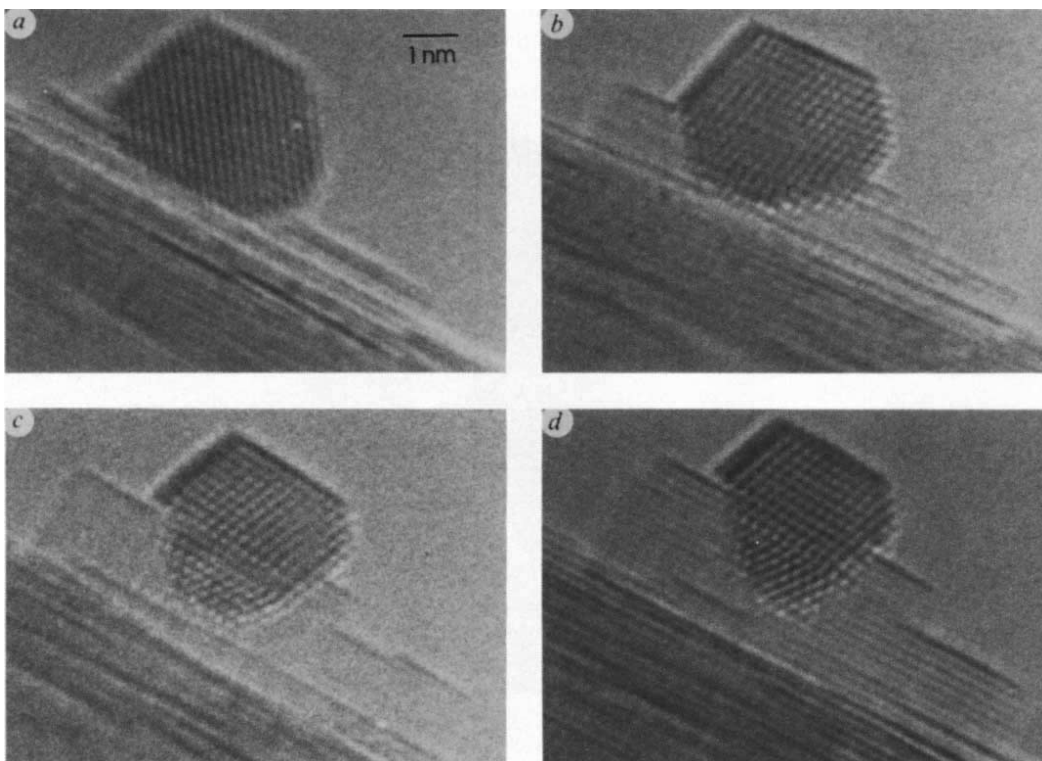


FIG. 3 Irradiation sequence of a gold particle on an MgO surface. *a*, Initial particle. *b*, After 10 min under low flux. *c*, After 30 min under low flux. Note that a twin boundary has formed in the lower part of the particle. *d*, After an additional 10 min of high-flux irradiation.

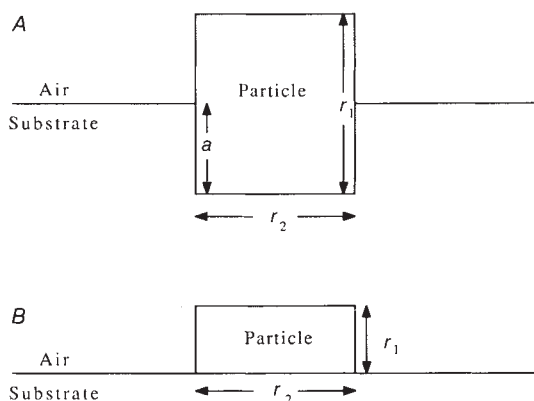


FIG. 4 A, Schematic diagram of a particle that is partly immersed in the substrate. B, A particle on a planar substrate. We assume for simplicity that the particle is of unit length normal to the figure and that the substrate surface is semi-infinite.

developed<sup>4-6</sup>, but only for planar interfaces. If the interface is not flat, the particle can either 'sink' into the substrate or 'float' on top.

Our experiments have studied the dispersion of microclusters of gold, prepared by reduction of organometallics<sup>7</sup>, onto magnesium oxide smoke particles (formed by burning magnesium ribbon in air). The specimens were imaged in a Hitachi H-9000 high-resolution electron microscope, operating at an accelerating voltage of 300 kV, which is capable of resolving individual atoms. The MgO particles, a few times 10 nm to a few micrometres in size, are cuboids with predominantly low-energy (100) facets, and the gold particles are found to decorate the surface of these cuboids (Fig. 1).

The experimental observations are shown in Figs 2 and 3. Figure 2 shows high-magnification images of a gold cluster on an MgO particle before and after being irradiated under low-electron-flux conditions ( $\sim 100 \text{ A cm}^{-2}$ ). On irradiation, the substrate engulfs the particle, or alternatively the particle sinks into the substrate. The lattice spacings of the overlayer that forms on the particle match the (200) spacings of the MgO substrate. Figure 3 shows a similar irradiation sequence of another particle, taken under atomic resolution, under low- and high-electron-flux conditions (the two fluxes differ by a factor of  $\sim 50$ ). In general it appears that sinking occurs faster at higher fluxes. Observations on many different particles confirm these results. In most cases a sinking particle orientates epitaxially to the engulfing substrate and can form defects, such as the twin boundary seen in Figs 3c and d. If the particle is epitaxial with respect to the substrate, sinking occurs much faster, whereas if there is some contamination on the particle the process may be retarded or even halted. Using the relation  $l = \sqrt{2Dt}$ , where  $l$  is the diffusion length and  $t$  the diffusion time, we obtain an estimate of the diffusion coefficient  $D$  for the encapsulation process of  $10^{-17} \text{ cm}^2 \text{ s}^{-1}$ , a value comparable to that found in SMSI (strong metal-substrate-interaction) systems<sup>8-10</sup>. We believe that here the electron irradiation, rather than thermal effects, provides the driving force for diffusion; this is consistent with the observation, during irradiation, of roughening of the MgO surface far from the gold clusters.

A simple model can be used to explain these phenomena. Figure 4A shows schematically a particle that is partially immersed in a substrate. Assuming that the strain energies at the interface are negligible compared with the surface energies (which holds if the interface is epitaxial) we can write the potential energy of the system as:

$$E = (r_2 + 2a)(\gamma_s^p + \gamma_s^{\text{sub}} + \gamma_1) + (2r_1 + r_2 - 2a)\gamma_s^p - r_2\gamma_s^{\text{sub}} \quad (1)$$

Here  $\gamma_s^p$  is the energy of the particle-vapour interface,  $\gamma_s^{\text{sub}}$  that

of the substrate-vapour interface, and  $\gamma_1$  the bonding energy of the interface; the distances  $r_1$ ,  $r_2$  and  $a$  are defined in Fig. 4A. By definition,  $\gamma_s^p$  and  $\gamma_s^{\text{sub}}$  are positive and  $\gamma_1$ , the energy released when the two surfaces are brought together, is negative. The change in total energy with respect to  $a$  is

$$\partial E / \partial a = 2(\gamma_s^{\text{sub}} + \gamma_1) \quad (2)$$

If  $\partial E / \partial a < 0$  (corresponding to a particle that forms a high-energy interface on a low-energy substrate facet, as may be the case here) the particle will sink; if  $\partial E / \partial a > 0$ , it will float. The system is thus thermodynamically unstable with respect to sinking or floating. Note that if the particle and the substrate consist of the same material the problem reduces to that of conventional sintering, and sinking becomes equivalent to heterogeneous sintering or Ostwald ripening. For a planar substrate surface (Fig. 4B), the energy balance is obtained similarly, and gives the criteria for wetting:

$$2\gamma_s^p + \gamma_1 < 0 \quad \text{Particle wets} \quad (3)$$

$$2\gamma_s^{\text{sub}} + \gamma_1 < 0 \quad \text{Substrate wets} \quad (4)$$

Note that the propensity of a particle to sink will be sensitive to changes in the surface free energies, as might be induced, for example, by chemisorption of impurities.  $\square$

Received 2 November 1988; accepted 31 January 1989.

1. Herring, C. *Structure and Properties of Solid Surfaces* (eds Gomer, R. & Smith, C. S.) 5-72 (Univ. of Chicago Press, 1953).
2. Rottman, C. & Wortis, M. *Phys. Rep.* **103**, 59-79 (1984).
3. Ajayan, P. M. & Marks, L. D. *Phys. Rev. Lett.* **60**, 585-587 (1988).
4. Winterbottom, W. L. *Acta Met.* **15**, 303-310 (1967).
5. Pillar, R. M. & Nuttig, J. *Phil. Mag.* **16**, 181-188 (1967).
6. Heyraud, J. C. & Metois, J. J. *Acta Met.* **28**, 1789-1797 (1980).
7. Teo, B. K. & Keating, K. J. *Am. chem. Soc.* **106**, 2224-2226 (1984).
8. Tauster, S. J. et al. *J. Am. chem. Soc.* **100**, 170-175 (1978).
9. Chen, Bor-Her. & White, J. M. *J. phys. Chem.* **86**, 3534-3541 (1982).
10. Spencer, M. S. *J. Catal.* **93**, 216-223 (1985).

ACKNOWLEDGEMENTS. We acknowledge funding by the NSF.

## Stress-induced Al-Cr zoning of spinel in deformed peridotites

Kazuhiro Ozawa

Geological Institute, Faculty of Science, University of Tokyo, Tokyo 113, Japan

DEFORMATION is one of the principal processes governing the microstructure of solid materials, but it may also affect their chemistries<sup>1-5</sup>. Chemical unmixing in non-hydrostatically stressed, initially homogeneous multicomponent solids has been predicted to occur during diffusion creep<sup>5,6</sup>. Yet there has been no report of chemical zoning induced by diffusion creep in either naturally or experimentally deformed solid-solution materials. Here I report Al-Cr zoning in elongated Cr-spinel ((Cr, Al, Fe<sup>3+</sup>)<sub>2</sub>(Mg, Fe<sup>2+</sup>)O<sub>4</sub>) grains from deformed peridotites and chromitites, which shows a consistent orientation of Al-rich and Al-poor regions with respect to the lineation and the shape of the spinel grain. I interpret this zoning as having been induced by deformation, and it is most reasonably explained by stress-directed lattice diffusion of Al and Cr.

I have determined the distribution of elements in spinel grains of spinel peridotites from the Miyamori ophiolitic complex and the Horoman ultramafic complex in north-east Japan, and in a chromitite from the Oman ophiolite, by electron-probe microanalysis (JEOL JCMS 733 Mk-II). In the analyses, the specimen stage was automatically driven at  $20\text{-}40 \mu\text{m s}^{-1}$  (50 ms counting time at each point and 1 or  $2 \mu\text{m}$  pixel size). The total number of pixels for each spinel analysis ranged from  $200 \times 200$  to  $300 \times 400$ .



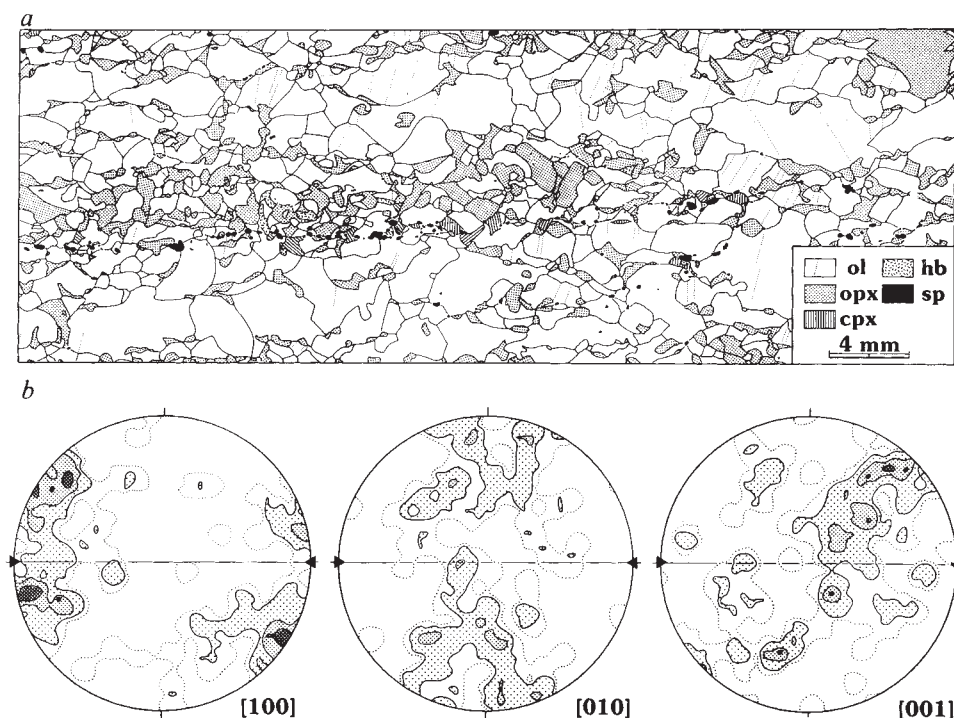


FIG. 1 *a*, Drawing of a thin section of a harzburgite sample from the Miyamori complex, cut normal to the foliation and parallel to the mineral aggregate lineation. ol: olivine; opx: orthopyroxene; cpx: clinopyroxene; hb: hornblende; sp: spinel. Dotted lines indicate sub-boundaries in olivine. *b*, Olivine preferred orientation in the harzburgite shown in *a*. Measurement on 100 grains from the olivine-rich portion; equal-area projection, lower hemisphere; contours denote 1, 2, 4 and 6% per 1% area. Arrowheads indicate the trace of lineation, and the horizontal line represents the trace of the foliation plane.

The harzburgites from the Miyamori complex show a high degree of mineral aggregate lineation defined by a linear arrangement of orthopyroxene and spinel grains (Fig. 1*a*). Foliation is generally weak and can be recognized by the flattened shape of olivine (Fig. 1*a*) and in the slightly elongated shape of spinel seen in the section cut perpendicular to the lineation (Fig. 2*b*). Olivine shows a preferred lattice orientation characterized by a [100] density maximum close to the lineation trend (Fig. 1*b*), which is commonly observed in deformed peridotites<sup>7</sup>. The harzburgite from the Horoman complex shows a porphyroclastic texture. Spinel and orthopyroxene porphyroclasts are stretched and pulled apart into smaller elongated grains, showing marked lineation. Neither plagioclase nor garnet is present in these samples. Petrofabric and textural studies on olivine in these peridotites clearly show that they have undergone high-temperature penetrative deformation (Fig. 1)<sup>8,9</sup>. The chromitite is of antinodular type and shows strong lineation defined by ellipsoidal olivine aggregates set in a spinel-rich matrix, suggesting penetrative deformation<sup>10</sup>. The spinel has a slightly elongated shape, with grain sizes ranging from 100 to 500  $\mu\text{m}$ .

Two thin sections, one perpendicular and one parallel to the lineation, were taken from a single spinel peridotite specimen of the Miyamori complex (Fig. 1) to search for three-dimensional chemical zoning. In the section parallel to the lineation, most of the spinels show an elliptical shape with their long axes lying parallel to the lineation (axial ratio  $<4:1$ , Fig. 1*a*). Some spinel

grains show only slight elongation, although their long axes remain consistently orientated. This is probably due to the effect of cutting on the spinel grains, whose shapes and orientations deviate from the ideal case in many instances. In the parallel section, aluminium is highly concentrated along the rims in the direction of the elongated axes, and is depleted along the rims in the direction of the short axes (Figs 2*a*, 3). Chromium shows a distribution pattern opposite to that of Al, whereas the ferric content does not show noticeable variation (Fig. 3). The Al-Cr zoning can be seen even with an optical microscope as a faint colour change: the colour of the rim in the direction of the long axis of elliptical spinel is, in most cases, slightly lighter than that of the rim in the direction of the short axis.

In the section cut perpendicular to the lineation, the spinel grains have a relatively circular or slightly elongated cross-section (axial ratio  $<2:1$ ). Aluminium is again highly concentrated at the rim in the elongated direction, and depleted at the rim in the direction of the short axis (Figs 2*b*, 3). The faint variation in colour observed for each spinel grain also demonstrates a consistency in the zoning polarity of these grains.

From microscopic observations of the element distributions in these two mutually perpendicular thin sections, we infer an idealized three-dimensional geometry of the spinel grains and their Al-Cr zoning as shown in Fig. 4. The spinel grain has a prolate ellipsoidal shape with its major axis parallel to the lineation. It shows multi-polar Al-Cr zoning, which is character-

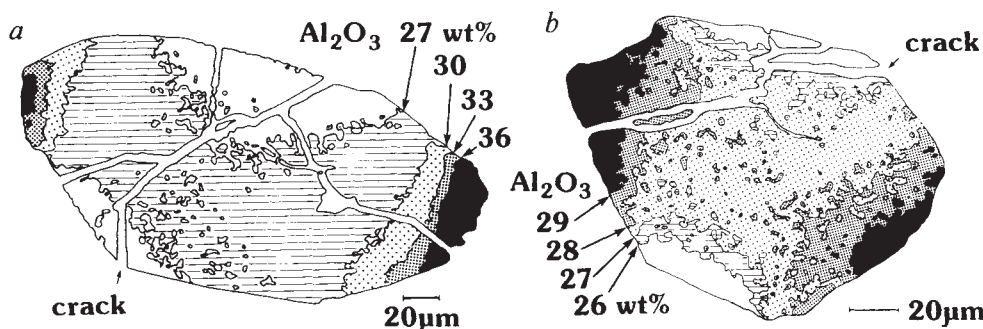


FIG. 2 Compositional zonings of chromian spinel in the Miyamori harzburgite (Fig. 1), measured on thin sections parallel to (*a*) and perpendicular to (*b*) the mineral lineation.

ized by (1) a maximum concentration of Al and minimum of Cr at the tips of the major axis, (2) a minimum concentration of Al and maximum of Cr roughly at the tips of the minor axis, and (3) an intermediate concentration of Al and Cr at the core. Multi-polar Al-Cr zoning in spinel grains is also observed in the other peridotites and chromitites studied (Fig. 2). The orientations, with respect to the lineation, of the Al-rich and Al-poor tips of the grains in these samples are consistent with those in the peridotite described here in detail.

This consistency of mineral lineation, preferred lattice orientations of olivine, shape of spinel, and polarity of Al-Cr zoning in the grains indicates that the multi-polar zoning is related to the deformation of the peridotites, and consequently of the spinel. Sub-solidus reactions (for example, Al-rich pyroxenes + olivine  $\rightarrow$  spinel + Al-poor pyroxenes) cannot be responsible for multi-polar zoning because symmetrical multi-polar zoning is found in spinel grains that are perfectly enclosed in or surrounded exclusively by olivine crystals. The coexistence of concentration maxima of both Al and Cr at the crystal boundary of the spinel clearly indicates that the original zoning pattern was not modified substantially after the deformation by any sub-solidus reactions caused through grain-boundary diffusion.

The multi-polar zoning does not represent an equilibrium distribution of the two elements resulting from the difference in size between substitutional atoms in non-hydrostatically stressed solids<sup>1,2</sup>, because the molar volume of Cr-spinel is greater than that of Al-spinel by  $\sim 10\%$  and so the polarity of zoning expected on this basis is quite the reverse of that observed. Furthermore, dislocation creep involving drag by a Cottrell atmosphere cannot be responsible for the zoning because solute atoms would be segregated around the moving dislocations<sup>11</sup>; aluminum is a solute atom in the Cr-rich portion, but not in the Al-rich portion (Fig. 3). Moreover, the net motion of dislocations from crystal faces in compression to faces in tension may be negligible, because in steady- or quasi-steady-state deformation mutual annihilation of dislocations of opposite signs, or sub-grain formation by alignment of edge dislocations of the same sign, may occur on a much smaller scale than the grain size<sup>7</sup>. A Cottrell atmosphere, consequently, cannot transport Al selectively from the Cr-rich faces to the Al-rich faces. A pressure-

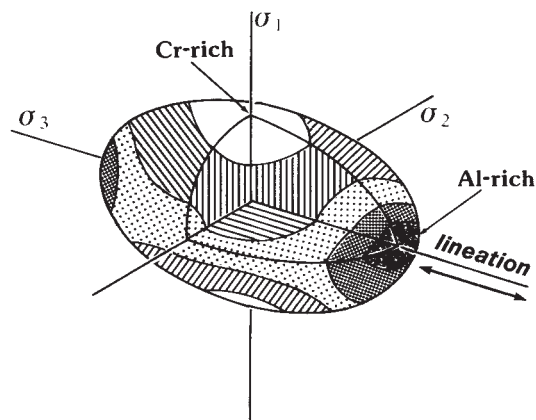


FIG. 4 Schematic picture of three-dimensional multi-polar Al-Cr zoning in spinel. Aluminium concentration decreases through densely stippled, sparsely stippled, and hatched to blank areas. The expected orientation of the three principal stress axes is also shown.

solution mechanism, with a silicate melt or a hydrous fluid phase along grain boundaries as transportation media, cannot have caused the multi-polar zoning because it develops even when the spinels are enclosed in olivine, which circumstance is unfavourable for such a mechanism.

The most probable deformation mechanism of spinel that can give rise to multi-polar zoning is diffusion creep (Nabarro-Herring or Coble creep<sup>12,13</sup>). In diffusion creep, vacancies are transported from crystal faces in tension to faces in compression, and matter flows in the opposite sense. Stress-induced chemical zoning (kinetic demixing) occurs when the diffusivities of the two end components are unequal and when the deformation is rate-limited by cation diffusion<sup>5,6</sup>. The faster ions are expected to be concentrated at crystal faces in tension and depleted at faces in compression.

Ando and Oishi<sup>14</sup> speculated that Al diffusivity in single-crystal  $\text{MgAl}_2\text{O}_4$  is as high as that of Mg and that both diffusivities are higher than that of oxygen. On the other hand, oxygen diffusion in polycrystalline  $\text{MgAl}_2\text{O}_4$  is as fast as Mg diffusion in single-crystal spinel<sup>15,16</sup>, suggesting that there is significant enhancement of oxygen diffusion along grain boundaries, as is found in  $\text{Al}_2\text{O}_3$  (ref. 17). These facts suggest that the oxygen diffusivity along grain boundaries could be of the same order of magnitude as the diffusivities of Al and Cr through the spinel lattice; thus the lattice diffusion of Al and Cr can be a controlling factor in the diffusion creep of spinel when the grain size is small enough.

Although no experimental data on diffusion coefficients for Al and Cr in Cr-spinel are available, the mobility of Al is probably greater than that of Cr, both because the melting temperature of Al-spinel is lower than that of Cr-spinel<sup>18-20</sup> and because the self-diffusion coefficient of Al in MgO is much greater than that of Cr (ref. 21). The observed Al-Cr zoning in deformed spinel is therefore consistent with kinetic demixing that is rate-limited by cation lattice diffusion.

The Al-Cr ratio for the cores of spinel grains is fairly constant in each of the peridotites and chromitites (Fig. 3). The multi-polar zoning is most pronounced when the core has an intermediate composition (Fig. 3). These facts suggest demixing of an originally homogeneous spinel<sup>3,6</sup>. The typically very low dislocation density (less than  $10^5 \text{ cm}^{-2}$ ) and the general lack of a preferred crystallographic orientation of spinel grains in peridotites and chromitites also support a diffusion-creep model<sup>22,23</sup>. The electron-channelling patterns and channelling contrast for the spinel grains demonstrate that they are single crystals with no sub-boundaries, except for the porphyroclasts in the

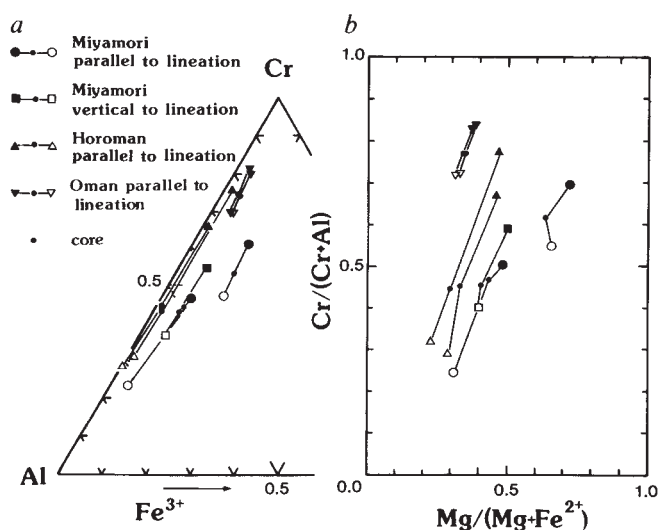


FIG. 3 Chemical composition of spinel, showing multi-polar zoning from the Miyamori and Horoman complexes and Oman ophiolite, plotted as *a*, the Cr-Al- $\text{Fe}^{3+}$  triangle, and *b*, the  $\text{Cr}/(\text{Cr}+\text{Al})$  versus  $\text{Mg}/(\text{Mg}+\text{Fe}^{2+})$  diagram. Solid symbols denote the rim along the short axis and open symbols the rim along the long axis. Analyses were made on thin sections cut parallel to the lineation, except for one Miyamori harzburgite which was cut perpendicular to the lineation.



Horoman peridotites<sup>24</sup>. The pole-figures for spinel in the chromitite from the Oman ophiolite, obtained with an X-ray texture goniometer, demonstrate that the spinel has no preferred lattice orientation. These facts indicate that dislocation creep is not a dominant mechanism in the deformation of spinel, and therefore, for those cases characterized by multi-polar zoning, deformation most probably results from the combination of cation transport (mainly through the spinel lattice (Nabarro-Herring creep)) and oxygen transport along interfaces (Coble creep).

This preliminary investigation of zoning in spinel grains and deformation fabrics in peridotites (ten samples) and chromitites (five samples) from several ophiolitic complexes shows that multi-polar Al-Cr zoning in spinel is not rare. The six peridotite and three chromitite samples, which are supposed to be deformed, contain spinel that exhibits such zoning. A statistical treatment of the orientation of concentration maxima and minima of the stress-induced Al-Cr zoning in these deformed spinel grains allows us to deduce the geometry of the principal stress axes. The orientation of the stress axes obtained independently from studying the olivine fabrics is of great value in determining the flow type acting within the upper mantle<sup>25</sup>. The maximum differences in element concentration in the spinel grains examined fall into distinct ranges, suggesting distinguishable differences in their deformation conditions. The distribution of Al-Cr concentration within each grain may therefore represent a potential piezometer under known temperatures, if the diffusivities of Al and Cr in spinel are known. Multi-polar zoning

of Al and Cr in spinel could constitute a useful tool in understanding upper-mantle deformation. □

Received 19 September 1988; accepted 26 January 1989.

1. Li, J. C. M., Nolfi, F. V. Jr & Johnson, C. A. *Acta metall.* **19**, 749-752 (1971).
2. Green, H. W. II, *Phil. Mag.* **41**, 637-647 (1980).
3. Kitamura, M., Matsuda, H. & Morimoto, N. *Proc. Jap. Acad.* **62**, 149-152 (1986).
4. Rutter, E. H. *J. geol. Soc. Lond.* **140**, 725-740 (1983).
5. Dimos, D. & Kohlstedt, D. L. *J. Am. Ceram. Soc.* **70**, 531-536 (1987).
6. Dimos, D., Wolfenstine, J. & Kohlstedt, D. L. *Acta metall.* **36**, 1543-1552 (1988).
7. Nicolas, A. & Poirier, J. P. *Crystalline Plasticity and Solid State Flow in Metamorphic Rocks* (Wiley, London, 1976).
8. Ozawa, K. *Lithos* **16**, 1-16 (1983).
9. Nida, K., *J. Jap. Ass. Miner. Petrol. Econ. Geol.* **70**, 265-285 (1975).
10. Cassard, D. et al. *Econ. Geol.* **76**, 805-831 (1981).
11. Hirth, J. P. & Lothe, J. *Theory of Dislocations* (McGraw-Hill, New York, 1968).
12. Herring, C. *J. appl. Phys.* **21**, 437-445 (1950).
13. Coble, R. L. *J. appl. Phys.* **34**, 1679-1682 (1963).
14. Ando, K. & Oishi, Y. *J. chem. Phys.* **61**, 625-629 (1974).
15. Oishi, Y. & Ando, K. *J. chem. Phys.* **63**, 376-378 (1975).
16. Lindner, R. & Akerström, A. *Z. phys. Chem. Neue Folge* **18**, 303-307 (1958).
17. Paladino, A. E. & Coble, R. L. *J. Am. Ceram. Soc.* **46**, 133-136 (1963).
18. Alper, A. M., McNally, R. N., Doman, R. C. & Keihl, F. G. *J. Am. Ceram. Soc.* **47**, 30-33 (1964).
19. Alper, A. M., McNally, R. N., Ribbe, P. G. & Doman, R. C. *J. Am. Ceram. Soc.* **45**, 263-268 (1962).
20. Lasaga, A. C. in *Reviews in Mineralogy* (eds Lasaga, A. C. & Kirkpatrick, N. J.) **8**, 261-319 (Miner. Soc. Am., 1981).
21. Whitney, W. P. II & Stüben, V. S. *J. Am. Ceram. Soc.* **54**, 348-352 (1971).
22. Doukhan, N., Doukhan, J.-C., Nicolas, A. & Secher, D. *Bull. Miner.* **107**, 777-793 (1984).
23. Christiansen, F. G. *Geol. Rundsch.* **74**, 61-76 (1985).
24. Christiansen, F. G. *Tectonophysics* **121**, 175-196 (1986).
25. Karato, S. in *High Pressure Research in Mineral Physics* (eds Manghni, M. H. & Syono, Y.) 455-471 (Terra Scientific Publ. Co., Tokyo, 1987).

ACKNOWLEDGEMENTS. I am grateful to Professor I. Kushiro and M. Toriumi and Mr G. P. Yumul for useful discussions and for a critical reading of the manuscript. I am also indebted to Dr S. Karato for his discussion, to Dr K. Kanagawa for assisting with X-ray texture goniometry and for advice, and to Drs S. Yanai and S. Umino for providing the Oman samples.

## Measuring thermal budgets of active volcanoes by satellite remote sensing

L. Glaze\*, P. W. Francis\* & D. A. Rothery†

\* Lunar and Planetary Institute, 3303 NASA Road 1, Houston, Texas 77058, USA

† Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, UK

ACTIVE volcanoes transfer both material and heat to the Earth's surface, but their thermal budgets are more difficult to monitor than their eruption rates, or the output of volatiles. Short-wavelength infrared data from the Landsat Thematic Mapper have been used to measure temperatures<sup>1,2</sup>, and the logical next step is to measure the total radiant energy flux,  $Q$ , in order to monitor changes in thermal budgets with time, and to allow the activity of different volcanoes to be compared. Here we report Thematic Mapper measurements of  $Q$  at Lascar volcano, north Chile, for December 1984 that are consistent with the earlier suggestion<sup>1</sup> that a lava lake was the source of the anomaly, and values for 1985-86 that are much lower, suggesting that fumarolic activity was then a more likely heat source. Our results show that satellite remote sensing may be used to monitor the activity of a volcano quantitatively, in a way not possible by conventional ground studies, and may provide a method for predicting eruptions.

To calculate  $Q$  we have used a two-wavelength technique<sup>2,3</sup> which calculates the temperatures and sizes of sub-pixel-sized radiant anomalies, and modified it to take account of background radiation. The radiation,  $L$ , detected by a sensor for a given pixel can be considered as being the sum of two components: the apparent background radiation,  $R_B$ , and the apparent anomalous radiation in each spectral band,  $R_5$  or  $R_7$  (corresponding to bands 5 and 7 of the Thematic Mapper, 1.55-1.75 and 2.08-2.35  $\mu\text{m}$  respectively). Each component is the product of the actual radiance of the hot area, and the proportion of the total surface area that it occupies. This condi-

tion can easily be solved for the anomalous radiance and inserted directly into the algorithm used in ref. 2, allowing the sub-pixel temperature to be calculated.

The choice of background temperature must be made very carefully, as the total energy flux increases as  $T^4$  (where  $T$  denotes temperature). Unfortunately, there are no data on the detailed structure of volcanic thermal anomalies. Where a convecting lava lake is present, the temperature/surface-area distribution could be modelled by a step function, with plateaux representing the temperatures of incandescent lava, chilled crust and background. For other kinds of anomalies, such as areas of fumarolic activity, an exponential or even a gaussian distribution may be more applicable. Given the data limitations, we have used temperatures derived from the margins of the anomaly to model the background temperature within the anomaly. Gen-

TABLE 1 Total radiant flux calculated using the two-band method for Lascar, Erta'Ale and Erebus

| Volcano   | Date                  | Data | $Q$<br>( $\times 10^7$ W) | Area<br>( $\text{m}^2$ ) | Flux density<br>( $\times 10^3$ W/ $\text{m}^2$ ) |
|-----------|-----------------------|------|---------------------------|--------------------------|---------------------------------------------------|
| Lascar    | Dec. 1984             | CC   | 6.7                       | 35,100                   | 1.9                                               |
|           | Mar. 1985             | CC   | 3.8                       | 20,700                   | 1.8                                               |
|           | July 1985             | CC   | 2.7                       | 18,000                   | 1.5                                               |
|           | Jan. 1986             | NN   | 1.9                       | 9,000                    | 2.1                                               |
|           | Aug. 1986             | NN   | 0.3                       | 3,600                    | 0.7                                               |
|           | Oct. 1986             | CC   | 2.6                       | 18,900                   | 1.4                                               |
|           | Nov. 1986             | A    | 0.9                       | 6,300                    | 1.4                                               |
|           | Nov. 1986             | NN   | 1.0                       | 7,200                    | 1.4                                               |
|           | Nov. 1987             | CC   | 4.7                       | 28,800                   | 1.6                                               |
|           | Nov. 1987             | NN   | 7.6                       | 34,200                   | 2.2                                               |
|           | Jan. 1988             | CC   | 5.8                       | 35,100                   | 1.6                                               |
| Erta'Ale  | Jan. 1986             | CC   | 6.9                       | 30,600                   | 2.3                                               |
| Niragongo | 1973 <sup>5</sup>     | —    | 11.3                      |                          |                                                   |
|           | 1959 <sup>10,11</sup> | —    | 54.0                      | 13,125                   | 41.1                                              |
|           | 1972 <sup>11</sup>    | —    | 1,220.0                   | 60,000                   | 203.3                                             |
| Kilauea   | ref. 10               |      | 93.0                      |                          |                                                   |
| Loki (Io) | ref. 12               |      | $126.0 \times 10^4$       | $4.1 \times 10^{10}$     | 0.3                                               |

Quantities measured directly for other volcanoes are also included. Average values of flux density are taken over the entire anomalous radiant area. A, NN and CC refer to uncorrected, nearest-neighbour and cubic-convolution resampled data types respectively.

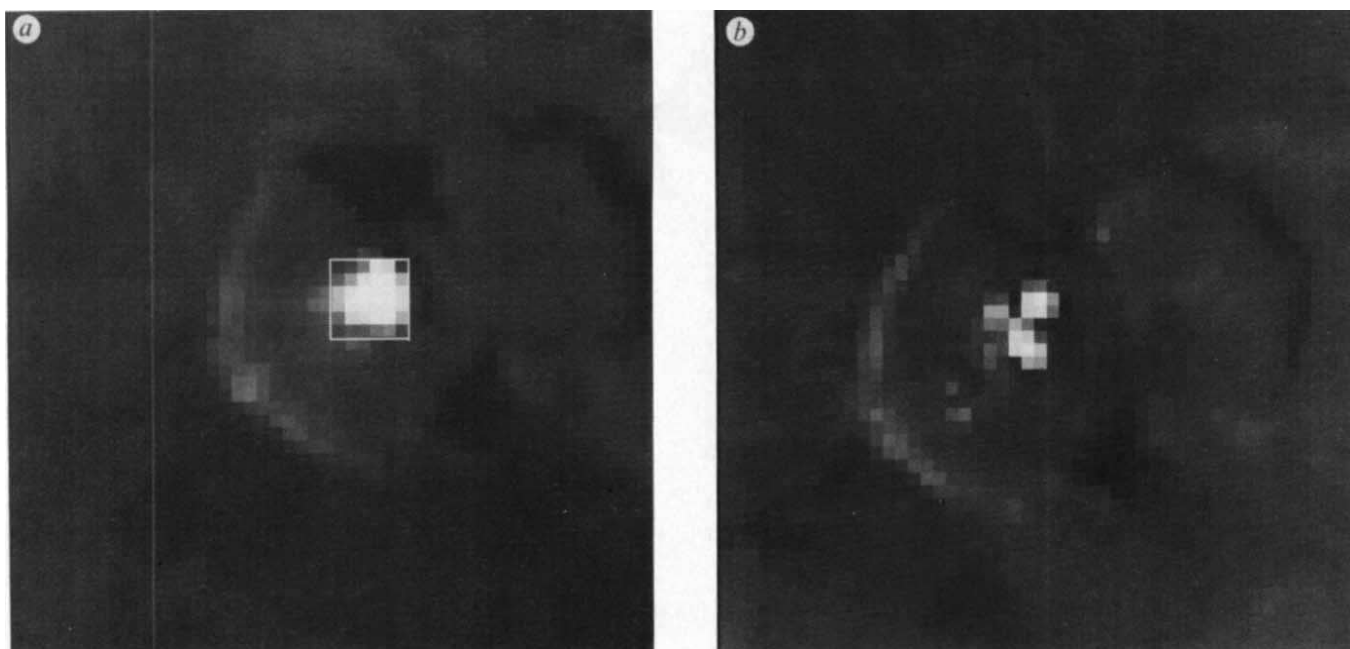


FIG. 1a, Landsat Thematic Mapper band 7 (2.08–2.35  $\mu\text{m}$ ) image showing radiant thermal anomaly in summit crater of Lascar volcano, 11 March 1985. Individual pixels are 30 m across. Brightest pixels indicate a thermal source with a temperature of 775  $^{\circ}\text{C}$  occupying 0.01 of the pixel. b, Band 7 image

of Lascar for 27 October 1986. Anomaly is highly fragmented after the eruption of 16 September 1986. Brightest pixel indicates thermal source at 1,100  $^{\circ}\text{C}$  occupying 0.001 of the pixel.

erally, along the margins of an anomaly there are some pixels that are thermally radiant in band 7 but not in band 5. A surface must be hotter than 100–150  $^{\circ}\text{C}$  over an entire pixel to be detectable in band 7, and hotter than 225–270  $^{\circ}\text{C}$  to be detectable in band 5. Thus, a perimeter pixel which is anomalous in band 7 but not in band 5 must have a pixel-integrated temperature in the range 100–225  $^{\circ}\text{C}$ . We assume that anomalous perimeter pixels have pixel-integrated temperatures at the lowest part of this range (100  $^{\circ}\text{C}$ ), and that this temperature prevails over the whole area of the anomaly, unless it is found, in calculating the sub-pixel temperatures of the perimeter pixels, that one or more pixels are radiant over the entire area of the pixel at some temperature greater than 270  $^{\circ}\text{C}$ . In this case, this higher temperature is used as the background temperature for radiance calculations for the pixels in the interior of the anomaly. The difference between the total  $Q$  calculated using background

values of 0  $^{\circ}\text{C}$  or 100  $^{\circ}\text{C}$  for anomalies greater than 15 pixels is less than 5%.

Some of the many sources of error in deriving temperature estimates from satellite data are discussed in ref. 2. One of the largest potential sources of error lies in the way that raw data are resampled for geometric correction. For TM data, uncorrected 'A' format images provide radiometrically accurate data but are impractical for most applications. 'P' format data are more practical, but standard 'cubic convolution' resampling algorithms cause any bright source to be smeared out. Nearest-neighbour interpolation retains individual pixel brightness values<sup>4</sup> but one out of every ten pixels is resampled twice. Because of external constraints, A-format, nearest-neighbour, and cubic-convolution data were all used in this study, but we recommend that A-format data should be used exclusively if possible. The magnitude of the error introduced by resampling

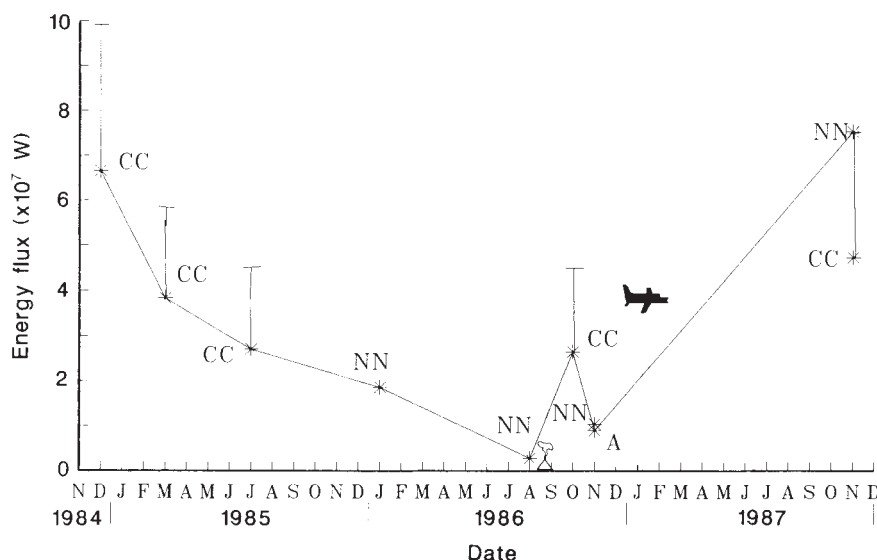


FIG. 2 Total flux radiated by Lascar over time. The volcano with plume represents the explosive eruption at Lascar of 16 September 1986. CC, NN and A denote the values which were calculated using cubic-convolution resampled, nearest-neighbour resampled, and A-format data respectively. Bars are shown which represent the error that occurs from using cubic-convolution as opposed to nearest-neighbour data. The airplane represents an aircraft observation overflight of the volcano made on 12 January 1987. Note the trend of decreasing flux before the eruption and the increase following.



decreases with the size of the anomaly and is unimportant for radiant anomalies larger than fifty pixels.

The only comparison available to us between ground-based and TM-based measurements of radiant thermal energy flux is provided by Erta'Ale, Ethiopia. In 1979, LeGuern *et al.*<sup>5</sup> reported that the summit caldera of Erta'Ale had contained two lava lakes between 1968 and 1974, and probably for very much longer previously. Our 1986 TM data showed that the two lava lakes still existed and were still very hot. The relative sizes of the two lakes had changed in the intervening twelve years, however, with the larger lake to the north becoming smaller and the smaller central lake becoming larger<sup>2</sup>. By summing all the radiant area, the minimum radiant thermal energy flux emitted by Erta'Ale was determined to be  $7 \times 10^7$  W (Table 1). This is very similar to the  $11.3 \times 10^7$  W measured directly by LeGuern in 1973. Although it is clear that changes have taken place at Erta'Ale since the ground data were acquired, the volcanic phenomena there seem overall to have been relatively consistent, and the close agreement between our TM data and the ground data suggests that the TM data are a reliable measure of radiant thermal energy. Because of the difficulties in obtaining synoptic measurements from the ground, satellite data may actually provide better estimates. We note, however, that basaltic lava lakes are subjected to rapid convective overturn and that they may change over short timescales. A single 'snapshot' observation of radiance should therefore be interpreted with caution<sup>6</sup>.

During a remote-sensing study of Central Andean volcanoes it was noted that a thermal anomaly existed in the crater of Lascar volcano, North Chile, in March and July 1985<sup>7</sup> (Fig. 1a). This was the only significant thermal anomaly to be discovered in the entire region, where about 50 major volcanoes should be regarded as 'active'<sup>8</sup>. The existence of this 'hot spot' implied that Lascar was in an unusually active condition, and this was subsequently demonstrated by an explosive eruption on 16 September 1986<sup>9</sup>. Since the eruption, several other TM images of Lascar have been acquired in order to monitor its activity both before and after the eruption (Fig. 1b). An anomaly was found to exist in each of the scenes acquired ranging from December 1984 to November 1987. The TM data provide two unique kinds of information on the activity of Lascar: an insight into the nature of the activity, and a record of the variations in  $Q$  with time. This latter parameter is also in a sense a measure of the overall level of activity of the volcano.

Figure 2 shows the variations in  $Q$  with time at Lascar. The average flux density (Table 1) is consistently lower than our calculated value for Erta'Ale as well as those directly measured for other lava lakes<sup>10,11</sup> although some of these values are questionable. Only in our first and last TM acquisitions (December 1984 and November 1987) were  $Q$  and the flux density at Lascar comparable to that from Erta'Ale. We suggest, on the basis of this and other evidence<sup>9</sup>, that a lava lake may have existed at some time, but that the persistent thermal anomaly at Lascar throughout 1985 and 1986 was due to an area of rocks in the summit crater being subjected to intense fumarolic heating.

Many more data are needed to constrain Lascar's activity. It is interesting, however, that the lowest recorded value ( $0.3 \times 10^7$  W) was measured one month before the major eruption of 16 September 1986. This reinforces a suggestion made by Glaze *et al.*<sup>9</sup> from petrographic considerations that the eruption was of phreatic origin and not related to the introduction of fresh magma into the volcano. It is also significant that the most recent data (November 1987) show a return towards previous high levels. We hope that future studies of Lascar and other volcanoes will provide us with a database that will enable us to understand better variations in radiant thermal energy and their relation to eruptive activity. Such a database may enable variation in  $Q$  to be used as a predictive tool. □

Received 10 October 1988; accepted 23 January 1989.

- Francis, P. W. & Rothery, D. A. *Geology* **15**, 614–617 (1987).
- Rothery, D. A., Francis, P. W. & Wood, C. A. *J. geophys. Res.* **93**, 7993–8008 (1988).
- Matson, M. & Dozier, J. *Photogramm. Engng. rem. Sens.* **47**, 1311–1318 (1981).
- Colwell, R. N. *Manual of Remote Sensing* (American Society of Photogrammetry, Falls Church, Virginia, 1983).
- Le Guern, F., Carbonnelle, J. & Tazief, H. *J. Volcan. geotherm. Res.* **6**, 27–48 (1979).
- Flynn, L. P., Mougins-Mark, P. J. & Gradie, J. C. *Science* (submitted).
- Francis, P. W. & McAllister, R. *Eos* **67**, 170–171 (1986).
- Francis, P. W. & de Silva, S. L. *Rem. Sens. Environ.* (in the press).
- Glaze, L. S. *et al. Bull. volcan.* (in the press).
- Bonnet, G. *Acad. R. Sci. Outre-Mer* **6**, 709–714 (1960).
- Le Guern, F. *J. Volcan. geotherm. Res.* **31**, 17–31 (1987).
- McEwen, A. S. *et al. J. geophys. Res.* **90**, 12345–12379 (1985).

ACKNOWLEDGEMENTS. This study was carried out at the Lunar and Planetary Institute, Houston, under the auspices of NASA. John Barker (NASA Goddard) provided valuable guidance on radiometric properties of TM data.

## Acute toxicity of aluminium to fish eliminated in silicon-rich acid waters

J. D. Birchall\*, C. Exley†, J. S. Chappell\* & M. J. Phillips†

\* ICI plc, PO Box 11, Runcorn, Cheshire WA7 4QE, UK

† Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, UK

**AN increased level of aluminium in acidified natural waters is a primary cause of fish death from damage to gill epithelia and loss of osmoregulatory capacity<sup>1–4</sup>. Aluminium toxicity depends on the species of aluminium present (cationic, neutral or anionic) and hence is affected by pH and the presence of complexing ligands, such as fluoride, and organic material, such as humic acid, which may ameliorate aluminium toxicity<sup>5,6</sup>. But silicic acid,  $\text{Si}(\text{OH})_4$ , present in natural waters as a consequence of the weathering of the aluminosilicates of rocks and soil minerals, has a strong and unique affinity for aluminium<sup>7</sup>, although its influence on toxicity has not been investigated. Here we show that, with an excess of Si over Al and with the formation of hydroxy-aluminosilicate species, the bioavailability of aluminium at pH 5 is reduced and acute toxicity is eliminated. Silicic acid concentration should therefore be considered as a key parameter in toxicity studies and could be important for the treatment of vulnerable waters.**

Atlantic salmon fry (*Salmo salar*; 1 g mass, post-first feeding) were maintained in tanks (120 fish per tank) with a through-flow of water at about pH 5 and containing various concentrations of aluminium and silicic acid. The concentration of aluminium in natural waters is in the range  $<1$  to  $>25 \mu\text{mol l}^{-1}$ , with acidic conditions and granitic catchment geology favouring the higher levels. Silicon concentration is within the range  $<20$  to  $>500 \mu\text{mol l}^{-1}$  and is usually low in acidic waters<sup>8,9</sup>. In our experiments, the aluminium concentration in all but the control experiment was at a level ( $6\text{--}7 \mu\text{mol l}^{-1}$ ) known to be acutely toxic at pH 5 (refs 4 and 6). All fish were acclimatized to control water at about pH 5 for one week before exposure to the water with high aluminium content. The conditions in each of the five experimental tanks are given in Table 1. Inorganic cationic aluminium species of low molecular weight are regarded as the toxic entity<sup>5</sup>. Here we define 'exchangeable aluminium' as the fraction of the total aluminium present that is retained on a sulphonate-functional cation-exchange resin (Amberlite IR 120). Fish were exposed for 96 h and the proportion of dead fish in each experimental tank was recorded at 12-h intervals (Fig. 1). The whole experiment was run three times and all the results presented have been amalgamated from the individual runs.

At a Si:Al ratio of 13 (tank 2), the acute toxicity of aluminium was eliminated and gill structures of the fish were normal when examined in the light microscope, in spite of the fact that this

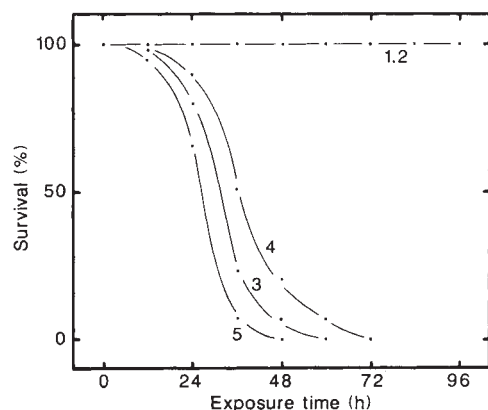


FIG. 1 Per cent survival of Atlantic salmon fry (1 g mass, post-first feeding) as a function of exposure time for the different treatment tanks (120 fish per tank). The curves are labelled for the respective tanks. All fish were kept under the control conditions for seven days before exposure and counts were made at 12-h intervals.

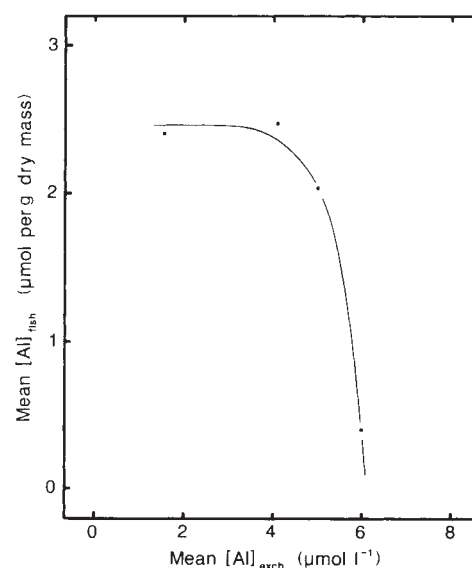


FIG. 2 Aluminium accumulation in fish (mean  $[Al]_{fish}$ ; see Table 2) as a function of exchangeable aluminium (mean  $[Al]_{exch}$ ) present in the treatment water (see Table 1).

water contained the highest level of exchangeable aluminium. Whole fish were collected at intervals of 12 hours and analysed for accumulated Al and Si. Table 2 shows the mean levels of aluminium and silicon recorded in fish from each of the experimental tanks. The fraction of the total aluminium that is exchangeable is increased with increasing silicic acid concentration (Table 1), so enhanced toxicity might be expected. Yet, as shown in Fig. 2, the accumulation of aluminium by fish falls sharply as the exchangeable aluminium increases.

This apparent paradox is resolved when we consider the interactions of hydroxy-aluminium species and silicic acid in dilute solutions. At pH 4 and above<sup>10</sup>, hydroxy-aluminium species and silicic acid interact to form hydroxy-aluminosilicate species which can be detected by infrared examination of the solids recovered by freeze-drying<sup>11</sup> or by the analysis of material retained on cation exchange resins, when it is found that aluminium and silicon are associated as species with Si:Al ratios of 0.25–0.5 (refs 12 and 13). These hydroxy-aluminosilicate species seem to be precursors of solid phases such as imogolite or protoimogolite<sup>11</sup>, and their growth to colloidal size is much slower than aluminium hydroxide. They persist as solution or subcolloidal species for a considerable time, as can be demonstrated by membrane filtration<sup>13</sup>. The resin (Amberlite IR 120) we use here to define exchangeable aluminium has a gel porosity excluding species of diameter  $>5$  nm. Driscoll<sup>14</sup> has pointed out that when aluminium solutions are passed through a cation-exchange resin column, there is competition for aluminium binding between the functional groups on the resin (sulphonate)

and ligands in solution. Aluminium is stripped from ligands such as sulphate and fluoride. In contrast, organic ligands such as citrate are stronger and aluminium is transported through an exchange column at pH 5. Citrate reduces the toxicity of aluminium<sup>5</sup>, presumably by binding aluminium more strongly than the anionic sites on gill epithelia. Competition for aluminium binding between gill sites and silicic acid in the bulk water seems to be the key to the reduced toxicity of aluminium that we have observed in this study.

Silicic acid and hydroxy-aluminium species start to combine at pH  $\geq 4$ , and the stability of the hydroxy-aluminosilicate species formed increases with increasing pH. The strength of the binding of silicate to hydroxy-aluminium exceeds that of sulphate above pH 5, of phosphate above pH 6 and of strong organic chelators such as citrate above pH 7 (refs 12 and 13). As such groups are present both on and within gill epithelia and mucus<sup>15</sup>, we suggest that the formation of hydroxy-aluminosilicate species (at  $\geq$  pH 4 and at  $>40 \mu\text{mol l}^{-1}$  silicic acid) blocks the binding of hydroxy-aluminium species to gill structures. Moreover, our results suggest that this effect is enhanced by increasing hydroxy-aluminosilicate stability in an alkaline boundary layer next to the gill surface<sup>15</sup>.

The interactions of silicic acid with hydroxy-aluminium species are unique<sup>7</sup> and as there are no related interactions with

TABLE 1 Conditions within the experimental treatment tanks

| Tank        | Mean pH | Total aluminium ( $\mu\text{mol l}^{-1}$ ) | Total silicon ( $\mu\text{mol l}^{-1}$ ) | Si:Al ratio | Exchangeable aluminium ( $\mu\text{mol l}^{-1}$ ) | Percentage of total aluminium exchangeable |
|-------------|---------|--------------------------------------------|------------------------------------------|-------------|---------------------------------------------------|--------------------------------------------|
| 1 (control) | 4.98    | 0.85 [0.07]*                               | 0.66 [0.10]                              |             | 0.18 [0.04]                                       | 21                                         |
| 2           | 4.96    | 7.15 [0.81]                                | 93.06 [8.01]                             | 13.0        | 6.00 [0.89]                                       | 83                                         |
| 3           | 4.93    | 6.70 [0.63]                                | 24.89 [0.82]                             | 3.7         | 5.00 [0.47]                                       | 74                                         |
| 4           | 4.94    | 6.22 [0.36]                                | 5.46 [0.23]                              | 0.9         | 4.11 [0.29]                                       | 66                                         |
| 5           | 4.94    | 6.26 [0.59]                                | 0.60 [0.06]                              | 0.1         | 1.52 [0.28]                                       | 24                                         |

The results are from three independent experimental runs with measurements recorded at 12-hour intervals. Mains water was used. It was aerated to remove residual chlorine, passed through a 5- $\mu\text{m}$ -pore-size filter to remove solids, through granular charcoal to remove organics (to  $<0.01 \text{ mg l}^{-1}$ ) and deionized to a conductivity of  $<0.05 \text{ ohm}^{-1} \text{ cm}^{-1}$ . The composition of the treatment water was adjusted to background levels for  $\text{Ca}^{2+}$  ( $50 \mu\text{mol l}^{-1}$ ),  $\text{Mg}^{2+}$  ( $10 \mu\text{mol l}^{-1}$ ),  $\text{Na}^{+}$  ( $<0.4 \text{ mmol l}^{-1}$ ) and  $\text{K}^{+}$  ( $30 \mu\text{mol l}^{-1}$ );  $\text{Cl}^{-}$ ,  $\text{F}^{-}$  and  $\text{SO}_4^{2-}$  were  $<0.5 \mu\text{mol l}^{-1}$ . The pH was maintained using  $\text{HNO}_3$  or  $\text{NaOH}$ . Aluminium and silicon levels were established using solutions of  $\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$  (BDH-Analar) and  $\text{Na}_4\text{SiO}_4$  (Alfa) as standards and graphite furnace atomic absorption spectroscopy for analysis. Temperature was controlled at  $15^{\circ} \pm 0.05^{\circ} \text{C}$ . Conditions in each tank were stabilized by operation for 48 h before introducing fish. Fish were acclimatized to the control tank (tank 1) for 1 week before introducing them to the other tanks.

\* Values are means with standard deviation shown in brackets.



TABLE 2 Aluminium and silicon levels in fish from experimental treatment tanks

| Tank        | Aluminium<br>( $\mu\text{mol per g dry mass}$ ) | Silicon<br>( $\mu\text{mol per g dry mass}$ ) |
|-------------|-------------------------------------------------|-----------------------------------------------|
| 1 (control) | 0.44 [0.22]*                                    | 0.01 [0.02]                                   |
| 2           | 0.40 [0.11]                                     | 0.54 [0.26]                                   |
| 3           | 2.04 [0.43]                                     | 0.35 [0.14]                                   |
| 4           | 2.49 [0.79]                                     | 0.33 [0.14]                                   |
| 5           | 2.38 [1.24]                                     | 0.08 [0.04]                                   |

The results are from three independent experimental runs. Three fish were removed from each tank every 12 h: moribund fish from tanks 3, 4 and 5, and healthy fish from tanks 1 and 2. These fish were lightly rinsed in dilute  $\text{HNO}_3$  to remove superficial aluminium on the skin before digestion in concentrated  $\text{HNO}_3$ . The rinse did not remove aluminium from the gill and served to reduce variability in the aluminium level measurement. Analysis for aluminium and silicon was by atomic absorption spectroscopy using a graphite furnace.

\* Values are means with standard deviation shown in brackets.

$\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  below pH 10, aluminium is removed from competition with these ions and with  $\text{Na}^+$  for binding at gill surfaces.

We conclude that the Si:Al balance in acid waters is a key, hitherto neglected, factor affecting the bioavailability and toxicity of aluminium. Increasing the level of silicic acid may be as significant as raising the pH in alleviating the toxic effect of

aluminium. A general question arises from this conclusion. Silicon has been claimed to be essential for the normal development of higher animals<sup>16,17</sup>; is silicon intrinsically an essential element for growth or does it merely limit the bioavailability of aluminium (in the mammalian gut, for example) as a result of the formation of hydroxy-aluminosilicates<sup>18</sup>? □

Received 7 November 1988; accepted 23 January 1989.

1. Driscoll, C. T. *Envir. Hlth Perspect.* **66**, 93–104 (1985).
2. Witters, H. E. *Aquat. Tox.* **8**, 197–210 (1986).
3. Muniz, I. P. & Leivestad, H. *Proc. Int. Conf. Impact Acid Precip.* (Norway. SNSF Project, 1980).
4. Schofield, C. L. & Trojnar, J. R. in *Polluted Rain* (eds. S. Toribara, T. Y., Miler, M. W. & Morrow, P. E.) 341–362 (Plenum, New York, 1980).
5. Driscoll, C. T., Baker, J. P., Bisogni, J. J. & Schofield, C. L. *Nature* **284**, 161–164 (1980).
6. Baker, J. P. & Schofield, C. L. *Wat. Air & Soil Pollut.* **18**, 189–309 (1982).
7. Iler, R. K. *The Chemistry of Silica* (Wiley, New York, 1979).
8. Farmer, V. C. *Ciba Fdn Symp.* **121**, 4–23 (1986).
9. Wright, R. F. et al. *Regional Surveys of Small Norwegian Lakes* (SNSF Project IR33/77, Oslo, 1977).
10. Lou, G. & Huang, P. M. *Nature* **335**, 625–627 (1988).
11. Farmer, V. C., Frazer, A. R. & Tait, J. M. *Geochim. cosmochim. Acta* **4**, 1417–1470 (1979).
12. Birchall, J. D. & Chappell, J. S. *Clin. Chem.* **34**, 265–267 (1988).
13. Chappell, J. S. & Birchall, J. D. *Inorg. chim. Acta* **153**, 1–4 (1988).
14. Driscoll, C. T. *Int. J. envir. Anal. Chem.* **16**, 267–283 (1984).
15. Exley, C. & Phillips, M. J. in *Recent Advances in Aquaculture* (eds Muir, J. F. & Roberts, J. R.) 225–341 (Helm, London, 1988).
16. Schwarz, K. & Milne, D. B. *Nature* **239**, 333–334 (1972).
17. Carlisle, E. M. *Science* **178**, 154–156 (1972).
18. Birchall, J. D. & Espie, A. W. *Ciba Fdn Symp.* **121**, 140–159 (1986).

ACKNOWLEDGEMENTS. We thank ICI Chemicals and Polymer Group for their support of this study and in particular for supporting the postgraduate work of C.E.

## Non-fisherian sex ratios with sex change and environmental sex determination

Eric L. Charnov\* & J. J. Bull†

\* Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

† Department of Zoology, University of Texas, Austin, Texas 78712, USA

FISHER'S theory of sex-ratio evolution<sup>1</sup> explains why 1:1 sex ratios are so prevalent<sup>2–8</sup>. Recent work has therefore emphasized situations in which biased sex ratios tend to evolve (for example, local mate competition<sup>7</sup>, non-mendelian inheritance of sex-ratio factors<sup>7,9,10</sup>, haplodiploid social insects<sup>11,12</sup> and others<sup>6,13</sup>). In species where individuals change sex during their adult lifetimes, and in species where sex is determined by environmental conditions experienced during pre-adult development, biased sex ratios are expected and frequently observed<sup>6</sup>. Sex-ratio models for particular kinds of sex change (SC) and environmental sex determination (ESD) have predicted excesses of the first sex<sup>6,14,15</sup> under SC, and of the sex developing in the 'worse' environment under ESD<sup>6,14–16</sup>. Here we generalize and unite these results in a single model that applies to a wide range of SC and ESD biologies. The principle that emerges from this model appears to explain strong female and male excesses seen in sex-changing fish and invertebrates and in nematodes with ESD. It is not known whether the female excess commonly seen in reptiles with ESD are also explained by this principle.

We develop this principle for sex change. Let the continuous variable  $t$  indicate an individual's age,  $\phi(t)$  the probability of surviving to age  $t$ ,  $m(t)$  the immediate fertility of a male at age  $t$ , and  $f(t)$  the immediate fertility of a female at age  $t$ . The lifetime male fertility [ $\chi(\tau)$ ] of individuals who become male first and change sex at age  $\tau$  is then

$$\chi(\tau) = \int_0^\tau m(t)\phi(t) dt \quad (1a)$$

and lifetime fertility as female is

$$y(\tau) = \int_\tau^\infty f(t)\phi(t) dt \quad (1b)$$

(ref. 6). The fertilities  $m(t)$  and  $f(t)$  may be thought of as relative fitnesses:  $m(t)$  is the average fitness of a male at age  $t$  divided by the fitness of an arbitrarily-chosen standard male, and similarly for  $f(t)$ ; because of this scaling,  $m(t)$  and  $f(t)$  as well as  $\chi(\tau)$  and  $y(\tau)$  do not measure absolute fitness because they do not depend on sex ratio. Note that, in the definitions of  $\chi(\tau)$  and  $y(\tau)$ , mortality and age/sex-specific fertility are functions of age but are not directly functions of  $\tau$ , the time of sex change. More complicated models for the evolution of sex change allow  $\phi(t)$ ,  $f(t)$  and  $m(t)$  to depend on  $\tau$  (refs 6, 17), but these additional complications do not alter the basic results from this simpler model, as will be presented here (see Fig. 1).

Models of sex-allocation theory typically consider the fate of a rare mutation adopting one sex-allocation strategy ( $\tau'$ ) in a population of individuals who otherwise adopt a different strategy ( $\tau$ ). The common genotype resists invasion provided that the net effect of the mutation on male plus female fitness is negative<sup>6</sup>, that is, if

$$\frac{\int_0^{\tau'} m(t)\phi(t) dt}{\int_0^\tau m(t)\phi(t) dt} + \frac{\int_\tau^\infty f(t)\phi(t) dt}{\int_\tau^\infty f(t)\phi(t) dt} < 2 \quad (2)$$

A common genotype whose sex-allocation strategy cannot be invaded by any alternative strategy is said to be evolutionarily stable<sup>2</sup>. The evolutionarily stable equilibrium for sex change must satisfy the principle that an individual's immediate fitness is independent of whether it is male or female<sup>6</sup>: if the population instead adopted a  $\tau$  for which fitness at age  $\tau$  was greater through one sex than the other, selection would favour altering  $\tau$  to extend the sex phase with higher reproductive success. To calculate the appropriate fitnesses, we may let female fitness at age  $t$  be proportional to fertility,  $c_1 f(t)$ . Male fitness will also be proportional to fertility, but in addition it must be scaled by the sex ratio: recalling Fisher's<sup>1</sup> sex-ratio argument, if adult males are half as common as adult females, the average number of children per male will be twice that per female. Male fertility must therefore be multiplied by  $c_1 \int_\tau^\infty f(t)\phi(t) dt / \int_0^\tau m(t)\phi(t) dt$

FIG. 1 A generalized sex-ratio theory for sex change. Assuming protandry (male first), male lifetime fertility [ $\chi(\tau)$ ] is plotted versus female lifetime fertility [ $y(\tau)$ ], for  $0 \leq \tau \leq \infty$ . Unlike in the text, here we allow the functions  $m(\cdot)$ ,  $f(\cdot)$ , and  $\phi(\cdot)$  to depend on  $\tau$  as well as on  $t$ . Such generalizations would apply, for example, if growth rate or death rate depended on sex. For protandry to be evolutionarily stable (uninvasible by protogyny or dioecy), this curve must be bowed outward, as drawn<sup>6,23</sup>. The evolutionarily stable value of  $\tau$  can be shown<sup>6</sup> to maximize  $\chi(\tau) \cdot y(\tau)$ . At this maximum,  $\partial y / \partial \chi = -y / \chi$ , or

$$\left| \frac{\partial y}{y} \right| = \left| \frac{\partial \chi}{\chi} \right|.$$

From this generalized version of equation (3), it follows that

$$\frac{r}{1-r} = \frac{\bar{F}/\partial y}{\bar{M}/\partial \chi}$$

The ratio of males to females [ $r/(1-r)$ ] is thus the inverse ratio of normalized average sex-specific fertilities. It is not obvious that this generalized sex-ratio result implies an excess of the first sex (at equilibrium), but this direction of sex-ratio excess is present in all of the cases we have studied, provided that male and female fertilities increase with age/size.

to yield male fitness. Equivalence of male and female fitness at the evolutionarily stable age of sex change,  $\hat{\tau}$ , therefore requires

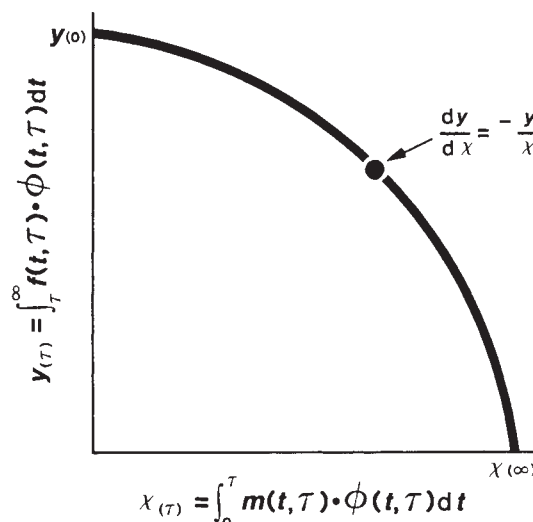
$$\frac{m(\hat{\tau})}{\int_0^{\hat{\tau}} m(t)\phi(t) dt} = \frac{f(\hat{\tau})}{\int_0^{\hat{\tau}} f(t)\phi(t) dt} \quad (3)$$

A simple relation between sex ratio and average fertility follows immediately from (3). The left-hand denominator may be written as  $k \cdot r \cdot \bar{M}$ , where  $k = \int_0^{\infty} \phi(t) dt$ ,  $r$  is the adult sex ratio,  $\int_0^{\infty} \phi(t) dt/k$ , and  $\bar{M}$  is the average male fertility,  $\int_0^{\infty} m(t)\phi(t) dt/k$ . The right-hand denominator may likewise be written as  $k \cdot (1-r) \cdot \bar{F}$ , where  $\bar{F}$  is the average female fertility. With this notation in mind, (3) may be written as

$$\frac{r}{1-r} = \frac{\bar{F}/f(\hat{\tau})}{\bar{M}/m(\hat{\tau})} \quad (4)$$

The evolutionarily stable adult sex ratio,  $r$ , is thus seen to depend on the average fertility of males relative to the fertility of a  $\hat{\tau}$ -age male, and on the average fertility of females relative to the fertility of a  $\hat{\tau}$ -age female. The adult sex ratio favours the sex experiencing the lower fertility values, where fertility is measured relative to a  $\hat{\tau}$ -age individual of the same sex. Of course, equation (4) applies to protogyny as well as protandry.

Result 4 is substantially more general than indicated by the present model. It applies whether age, size or some other combi-



nation of traits is used to trigger sex change. More generally, the form of equation 4 applies even if male and female fertilities or mortality depend on sexual experience (developed in Fig. 1). Result 4 is only approximately true if individuals suffer a temporary decrease in fertility during sexual transition<sup>17-19</sup>, but the error is trivial if the decrease in fertility applies to only a brief part of the second sexual phase. Result 4 is not robust to high levels of mortality incurred during sexual transition, although mortality immediately after sexual transition is incorporated. There is, however, no evidence for any high mortality cost during sex transition, at least in fish<sup>17-19</sup>. Finally, result 4 applies to environmental sex determination, with sex determined in response to the environmental cue  $t$  either by the parent<sup>20,21</sup> or by the offspring<sup>6,14,15,20</sup>;  $m(t)$  is then the expected lifetime fertility of a male raised at  $t$ ,  $f(t)$  is the expected lifetime fertility of a female, and  $\hat{\tau}$  defines the conditions for which a male and a female have the same expected lifetime fitness. In this latter application,  $r$  is the evolutionarily stable primary sex ratio. Thus, this theory applies to the Trivers-Willard model of parent-controlled ESD<sup>20</sup>.

Adult sex ratios estimated for protogynous fish and protandrous invertebrates are typically skewed, sometimes toward males and other times toward females. However, sex change occurs in two basic forms: protandry (male function first in life) and protogyny (female first). From sex-allocation theory, protandry is evolutionarily stable if female fertility increases with

TABLE 1 Sex ratios in sequential hermaphrodites

| Taxonomic group and location                                   | Proportion male       | Ref.   |
|----------------------------------------------------------------|-----------------------|--------|
| <b>Protandry (male first)</b>                                  |                       |        |
| Pandalid shrimp (several species, Northern Hemisphere)*        | 0.55-0.75             | 23, 24 |
| <i>Patella vulgata</i> (limpet, UK)                            | 0.75                  | 25     |
| <i>Crepidula</i> (limpet, North America)†                      |                       |        |
| <i>C. fornicata</i> (6 locations or years)                     | 0.4 (1), 0.5-0.75 (5) | 26     |
| <i>C. convexa</i> (3 locations)                                | 0.35-0.6              | 26     |
| <i>Amphiprion</i> (anemonefish, several species and locations) | 0.5                   | 33, 34 |
| <b>Protogyny (female first)</b>                                |                       |        |
| Scaridae (Parrotfish 3 sp., Caribbean)‡                        | 0.15, 0.23, 0.26      | 27     |
| (3 species, E. Australia)‡                                     | 0.21, 0.33, 0.38      | 28     |
| Labridae (wrasses 2 sp., Caribbean)‡                           | 0.11, 0.24            | 29     |
| <i>Labroides</i> (cleaner wrasse, Australia)                   | 0.15                  | 30     |
| <i>Epinephelus</i> (red grouper, Florida)                      | 0.17                  | 31     |
| <i>Anthias</i> (many locations)                                | 0.1 or less           | 32     |
| <i>Centropyge</i> (anglefish, Japan)                           | 0.29                  | 35     |

\* Estimates were obtained from adult mortality and sex-change data.

† Both species show a positive correlation between sex ratio and population density.

‡ Some species exhibit a pure male morph as well as the sequential hermaphroditic morph; sex-ratio data were included here only if at least 98% of the population was of the hermaphroditic morph.



age faster than does male fertility, that is, if the ratio  $f(t)/m(t)$  increases monotonically with age; protogyny is evolutionarily stable if this ratio decreases<sup>6,17,18</sup>. Female fertility may often be determined by fecundity and should increase with age. Male fertility may also usually increase with age, but the extent of this increase will depend on the social system. The greatest potential for an increase in  $m(t)$  with  $t$  should be associated with extreme polygyny, hence we expect that protogyny will be associated with polygyny. Protandry, on the other hand, should be found in species in which single males do not monopolize many matings.

If fertility increases with age for both sexes, a protandrous male should be most fit just before changing sex whereas a protandrous female is least fit just after changing sex, so

$$\frac{\bar{M}}{m(\tau)} < 1 < \frac{\bar{F}}{f(\tau)} \quad (5)$$

From results 4 and 5 we expect a male excess under protandry, and a male excess is nearly always observed (Table 1). Under protogyny, these inequalities should be reversed, and the expected female excess is also supported by the data. The obvious alternative hypothesis—that the second sex is rarer because the brief period of sexual transition involves high mortality—finds no support in observations on sex change in fish<sup>17–19</sup>.

Several of the protogynous fish listed in Table 1 exhibit social systems in which a few males dominate a social group and appear to prevent females from changing sex. Removal of a male leads to rapid sex change of the largest or most dominant female<sup>19</sup>. Such systems are easily incorporated into the foregoing framework, except that fertility is a discrete rather than continuous function of group size. Equality 3 then becomes an inequality, and the sex-ratio skew will be even greater than predicted by equation 4. Anemonefish (whose observed sex ratio is 0.5) provide a special case because the social system limits group size to two individuals, in which case male fertility depends entirely on group composition, not age. The limpet *Crepidula convexa* also violates the usual protandry sex-ratio pattern; the reason for this exception is unexplained but may stem from the mobility of males, which may offer an advantage to small size [ $m(t)$  would decrease with  $t$ ].

Sex ratios under environmental sex determination are also skewed from 0.5, favouring males in a mermithid nematode<sup>22</sup> and females in reptiles (data were summarized in ref. 8). A mermithid's sex is determined in response to nutrition during its juvenile phase when it parasitizes an insect: a worm developing in an undernourished, crowded, and/or small host becomes a male, whereas a worm developing in a large, well-nourished host becomes a female<sup>6,22</sup>. The mermithid case thus parallels protandry. The primary sex ratio lies in the range 0.7–0.9 for each of five populations of the mosquito parasite *Romanomermis nielsenii*, which is remarkable in view of the fact that the distributions of  $t$  differ greatly among populations<sup>22</sup>.

Data from a variety of turtles and crocodilians with temperature-dependent sex determination suggest that primary sex ratios are often as extreme as 0.3, although the sex ratio in some species is consistent with 0.5 (ref. 8). These skewed sex ratios fall within the range observed in sex-changing fish and mermithids, but it is not yet plausible that incubation temperature has such large effects on fertility<sup>21</sup>.

The present theory offers the best explanation for overall trends in the data in that the direction of observed sex-ratio excesses are predicted (except in *Crepidula*). The observed sex-ratio excesses are often extreme, particularly under protogyny and in the nematodes. The present theory can potentially account for the observed magnitudes of deviations from 0.5, but quantitative predictions await studies of the actual fertility functions,  $m(t)$  and  $f(t)$ . Perhaps the most interesting aspect of this theory as applied to sex change is that the generalizations apply to the adult or breeding sex ratio. Under typical fisherian sex-ratio selection for dioecious species, it is the primary sex ratio that

is predicted to obey a generalization (equal investment), and the adult sex ratio is simply a consequence of the prevailing sex-specific mortality rates. □

Received 21 November 1988; accepted 31 January 1989.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
3. Shaw, R. F. & Mohler, J. D. *Am. Nat.* **87**, 337–342 (1953).
4. Williams, G. C. *Adaptation and Natural Selection* (Princeton Univ. Press, Princeton, 1966).
5. Leigh, E. G. *Oxford Surveys evol. Biol.* **3**, 187–223 (1986).
6. Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, Princeton, 1982).
7. Hamilton, W. D. *Science* **156**, 477–488 (1967).
8. Bull, J. J. & Charnov, E. L. *Oxford Surveys evol. Biol.* **5**, 96–135 (1988).
9. Lewis, D. *New Phytologist* **40**, 56–63 (1941).
10. Howard, H. W. *J. Genet.* **44**, 143–159 (1942).
11. Hamilton, W. D. A. *Rev. Ecol. Syst.* **3**, 193–232 (1972).
12. Trivers, R. L. & Hare, H. H. *Science* **191**, 249–263 (1976).
13. Karlin, S. & Lessard, S. *Theoretical Studies on Sex Ratio Evolution* (Princeton Univ. Press, Princeton, 1986).
14. Frank, S. A. & Swingland, I. S. *J. theor. Biol.* **135**, 415–418 (1988).
15. Charnov, E. L. & Bull, J. J. *J. theor. Biol.* (in the press).
16. Bull, J. J. *Heredity* **46**, 9–26 (1981).
17. Warner, R. R. *Environmental Biology of Fishes* Vol. 22, 81–90 (1988).
18. Warner, R. R. *Trends Ecol. Evol.* **3**, 133–136 (1988).
19. Shapiro, D. Y. *Science* **209**, 1136–1137 (1980).
20. Trivers, R. L. & Willard, D. E. *Science* **179**, 90–92 (1973).
21. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, Menlo Park, 1983).
22. Blackmore, M. S. thesis, Univ. Utah (1989).
23. Charnov, E. L. *Am. Nat.* **113**, 715–734 (1979).
24. Charnov, E. L., Gotshall, D. & Robinson, J. *Science* **200**, 204–206 (1978).
25. Ballantine, W. J. thesis, London Univ. (1961).
26. Hoagland, K. E. *Malacologia* **17**, 365–391 (1978).
27. Robertson, D. R. & Warner, R. R. *Smithsonian Contrib. Zool.* **255**, 1–26 (1978).
28. Choat, J. H. & Robertson, D. R. in *Intersexuality in the Animal Kingdom* (ed. Reinboth, R.) 263–283 (Springer, Berlin, 1975).
29. Warner, R. R. & Robertson, D. R. *Smithsonian Contrib. Zool.* **254**, 1–27 (1978).
30. Robertson, D. R. & Choat, J. H. *Proc. Second Int. Symp. Coral Reefs 1* (Biology), 217–225 (1974).
31. Moe, M. A. *Florida Dept. Nat. Resources Prof. Paper* **10**, 1–95 (1969).
32. Shapiro, D. Y. & Lubbock, R. J. *J. theor. Biol.* **82**, 411–426 (1980).
33. Fricke, H. & Fricke, S. *Nature* **266**, 830–832 (1977).
34. Fricke, H. W. *Z. Tierpsychol.* **50**, 313–326 (1979).
35. Moyer, J. T. & Nakazono, A. *Jap. J. Ich.* **25**, 25–39 (1978).

ACKNOWLEDGEMENTS. We thank J. Seger for improving the clarity of the text.

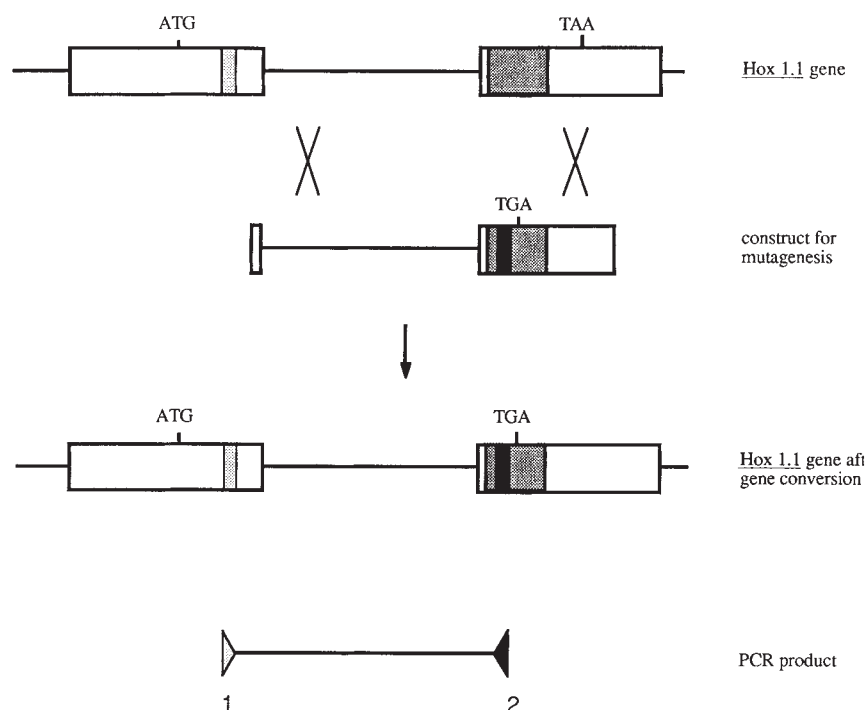
## Production of chimaeric mice containing embryonic stem (ES) cells carrying a homoeobox *Hox 1.1* allele mutated by homologous recombination

Andreas Zimmer & Peter Gruss

Max-Planck Institut für biophysikalische Chemie,  
Abt. Molekulare Zellbiologie, Am Fassberg, 3400 Göttingen, FRG

SEVERAL mouse gene families related to *Drosophila* developmental control genes and containing a homoeobox, a paired box or a finger domain, have been cloned and structurally analysed. On the basis of structural similarities to the *Drosophila* genes and of their spatially and temporally restricted expression patterns during mouse embryogenesis, it has been proposed that these mammalian genes also are involved in the control of development<sup>1–4</sup>. To elucidate the function of homoeobox genes by genetic means, mouse mutants must be generated. We have developed a technique for mutagenesis *in vivo* and have used it to mutate the homoeobox *Hox 1.1* gene. *In vivo* mutagenesis was achieved through homologous recombination between an endogenous *Hox 1.1* allele and a microinjected mutated gene in pluripotent embryonic stem (ES) cells<sup>5–9</sup>. Mutant cells were identified by means of the polymerase chain reaction (PCR)<sup>10</sup> and mutant clones were used to generate chimaeric mice. Because the homologous recombination event is formally a gene conversion event and no selection is required to screen for cells carrying the mutated allele, *in vivo* mutagenesis allows specific alterations in the target sequence to be made without the introduction of any other sequences.

FIG. 1 Strategy for the mutagenesis of the *Hox 1.1* gene and its detection by PCR. An *FspI* fragment of the *Hox 1.1* gene (nucleotides 367–1937 relative to the translational start codon, see Fig. 3) was cloned into Bluescript (Stratagene) and a 20-bp oligonucleotide inserted into the *EcoRI* site of the homoeobox. Homologous recombination of this fragment with the endogenous *Hox 1.1* gene in the recipient cells introduced this oligonucleotide into the endogenous gene. Consequently, transcripts from this allele had a frameshift mutation with a stop codon directly downstream of the oligonucleotide in-frame. To screen the recipient cells for homologous recombination events by PCR, we synthesized two primers, of which the first (Primer 1; TTC CGC ATC TCA CCC TGG AT) was specific for the first exon of *Hox 1.1*, binding 5' to the *FspI* site, and the second (Primer 2; AAT TGT GAG GTA CCG CTG AC) was identical to the inserted oligonucleotide. As the normal *Hox 1.1* allele contained no binding site for primer 2 and the microinjected fragment contained no binding site for primer 1, only the mutant allele can contain both priming sites. The homoeobox is indicated by a dark-shaded box. Primer 1 and its binding site are depicted as a light-shaded arrowhead and box, respectively. Primer 2 and the oligonucleotide are shown as a black arrowhead and box, respectively.



Relatively few genes are amenable to screening for homologous recombination using selective conditions<sup>11,12</sup>. We therefore designed a general strategy for targeting specific mutations to genes avoiding the need for selection. We used microinjection of DNA because it is the most efficient method for introducing DNA into cells<sup>13</sup> and increases the ratio of homologous to illegitimate recombination (M. R. Capecchi, personal communication). We limited the amount of non-homologous DNA to the introduced mutation and we used the PCR to screen cell pools for homologous recombination events. We used this procedure to interrupt the homoeobox of a *Hox 1.1* allele with a 20-base pair (bp) oligonucleotide. The protein encoded by the resulting mutated allele has an intact amino-terminus and a truncated carboxy-terminus and consequently lacks the putative DNA binding domain. The oligonucleotide sequence used not only ensured the introduction of a *KpnI* site at the expense of the *EcoRI* site (see Fig. 3d) but also served as a priming site for the amplification of the *Hox 1.1* intron by the PCR technique.

To obtain substrate for homologous recombination, the intron and second exon of the *Hox 1.1* gene<sup>15,16</sup> were subcloned and a 20-bp oligonucleotide was inserted into the *EcoRI* site of the homoeobox (Fig. 1). To test the efficiency of homologous recombination and the specificity of the PCR analysis, the *FspI*-fragment (see Figs 1 and 3) was first injected into large pools (250–600 cells) of 3T3 and P19 cells. As outlined in Fig. 1, the PCR with genomic DNAs using primers 1 and 2 (arrowheads in Fig. 1) will amplify a 1.1-kilobase (kb) DNA fragment including the intron–exon boundaries of *Hox 1.1* only if the microinjected fragment has correctly recombined with the genomic DNA. Figure 2a shows the amplification of the predicted 1.1-kb band from two of the microinjected pools (lane 1 and 2), compared with the negative signal from a third pool (lane 3). The plasmid pH1.1/O-BF (Fig. 2c), which includes the entire *Hox 1.1* gene and the inserted oligonucleotide, served as a positive control (lane 4). A total of 3,180 cells in seven pools were microinjected and at least five independent homologous recombination events were detected. The frequency of homologous recombination was therefore higher than 1 in 640 cells, and individual events could be detected among large cell pools.

Having established the reliability of the PCR analysis of homologous recombination, we injected the *FspI* fragment into

pools of 50–200 ES cells growing in 6-cm tissue culture dishes. The ES cell line D3 (ref. 7) was used because it contributes to the germ line of chimaeric animals and maintains this capability after genetic manipulations<sup>9</sup>. Figure 2b shows the PCR screening of 12 microinjection experiments. Southern blot analysis of the amplified DNA using an intron-specific probe (see Fig. 2c) detected a 1.1-kb fragment in five lanes, corresponding to at least five independent recombination events. Table 1 summarizes the analysis of the microinjected ES cell pools. The average frequency of homologous recombination was 1 event per 150 microinjected cells.

Two ES cell pools (1 and 7; Fig. 2b and Table 1) were chosen for further analysis. Single cell clones from these pools were obtained by picking individual cells with a glass capillary<sup>17</sup>. DNA was prepared from individual clonal cell lines, digested with *KpnI* and used for Southern blots. Figure 3a shows the result of one subcloning experiment in which four individual clones were examined. Because the inserted oligonucleotide has a *KpnI* restriction enzyme site, the appearance of a new 5.4-kb band in homologously recombined clones, in addition to the 14-kb band from the normal allele, was expected after hybridization with a probe specific for the first exon (probe 1 in Fig. 3d). The presence of this band is shown in lane 1 of Fig. 3a. The ES cell DNA became contaminated with feeder cell DNA

TABLE 1 Analysis of microinjected ES cell pools by PCR

| Pool no. | No. of injected cells | PCR |
|----------|-----------------------|-----|
| 1        | 50                    | +   |
| 2        | 71                    | +   |
| 3        | 54                    | +   |
| 4        | 45                    | –   |
| 5        | 108                   | –   |
| 6        | 130                   | –   |
| 7        | 50                    | +   |
| 8        | 48                    | –   |
| 9        | 42                    | –   |
| 10       | 46                    | –   |
| 11       | 50                    | –   |
| 12       | 25                    | +   |

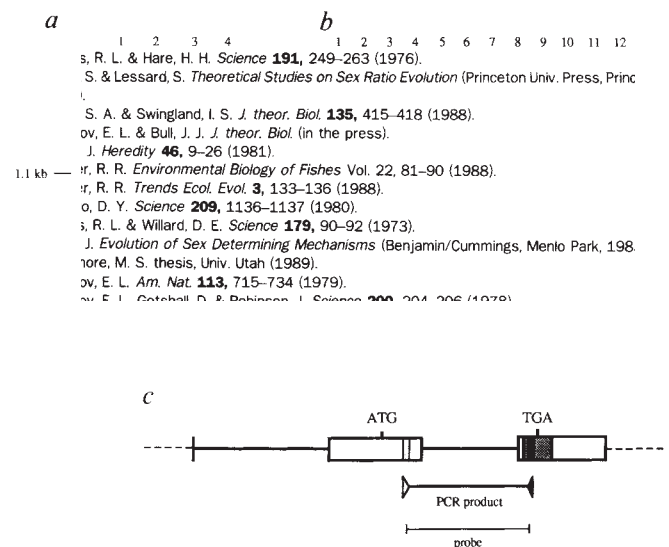
Details of the analysis are given in Fig. 3b.

because the ES cells were not allowed to form large colonies during subcloning. Thus, the 14-kb *Hox 1.1* fragment represents not only the one wild-type allele from the ES cells, but also both wild-type alleles from the feeder layer cells. In total, 161 individual clones from pool 1 and 165 clones from pool 7 were analysed by Southern blotting; one clone from pool 1 and two clones from pool 7 contained a mutated *Hox 1.1* allele. To confirm and extend the Southern analysis, clones from both pools were cultured until large ES cell colonies formed on the feeder cells. The relative amount of ES cell DNA was then much

higher as demonstrated in Fig. 3b and c. A Southern blot of *KpnI* digested DNA from a clonal line derived from pool 1 was hybridized to probe 1 (Fig. 3b, lane 1) and probe 2 (Fig. 3b, lane 2) (see Fig. 3c for details of the probes). The relative intensities of the 5.4-kb and the 8.6-kb bands, corresponding to the 5' and 3' fragments, respectively (see Fig. 3d), compared with that of the 14-kb fragment, show that at least 50% of the ES cells carry the mutated allele. DNA of a clonal line from pool 7 was digested with *StuI/KpnI* (Fig. 3c, lane 1) and *KpnI* (Fig. 3c, lane 2) and Southern blots were subsequently hybrid-

**FIG. 2 a**, Southern analysis of PCR-amplified genomic DNA from microinjected 3T3 cell pools (lanes 1–3, for details see text) and plasmid pH1.1/O-BF (lane 4). **b**, Southern analysis of PCR-amplified genomic DNA from microinjected D3 cell pools. **c**, Schematic representation of plasmid pH1.1/O-BF. A 3,643-bp 5' *BamHI*-*FspI* 3' fragment of the *Hox 1.1* gene (nucleotides –1,711 to +1,932 relative to the translational start codon) was cloned into Bluescript (Stratagene) and the disrupting oligonucleotide was cloned into the *EcoRI* site of the homoeobox.

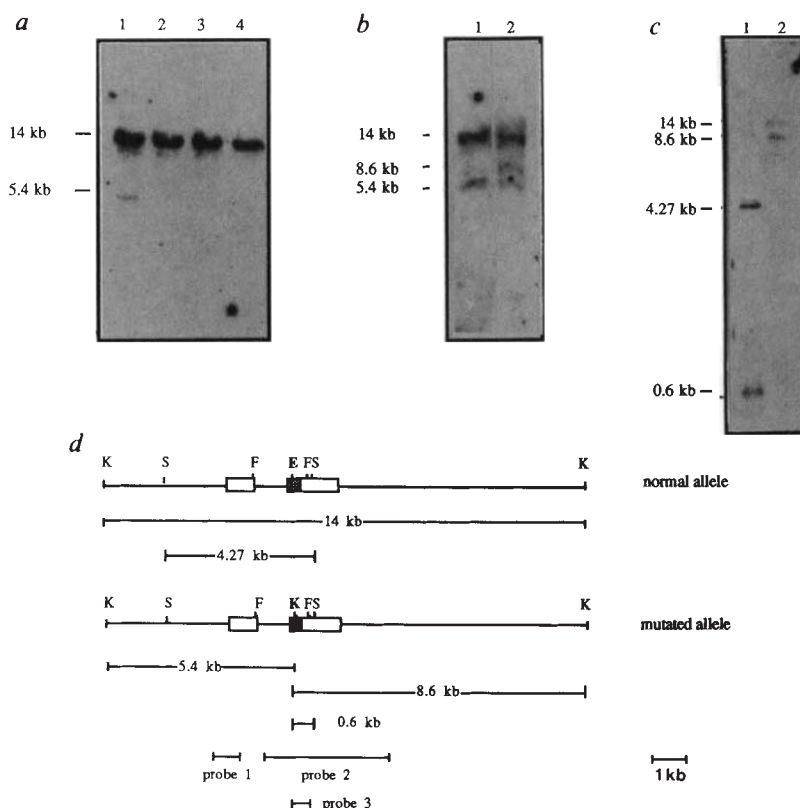
**METHODS.** The construct outlined in Fig. 1 was microinjected into the nucleus<sup>13</sup> of 3T3 or D3 cells (3T3: 200–600 cells; D3: 50–200 cells) at a concentration of 300 ng ml<sup>-1</sup>, corresponding to approximately five DNA molecules per cell with an injection volume of 10<sup>-11</sup> ml. The cells were grown for 4–7 days after microinjection, treated with trypsin and replated using the same dish. The cells were collected again after 4–7 days. Half of these were used for DNA analysis and the rest stored in liquid nitrogen. Genomic DNA or control plasmid (1 µg and 15 µg, respectively) was subjected to 30 PCR cycles using heat-stable *Taq*-polymerase<sup>14</sup> in a 50 µl reaction mixture containing 10% dimethyl sulphoxide, 67 mM Tris-HCl (pH 8.8), 16.6 mM NH<sub>4</sub>SO<sub>4</sub>, 6.7 mM MgCl<sub>2</sub>, 10 mM 2-mercaptoethanol, 170 µg ml<sup>-1</sup> bovine serum albumin (BSA), 45 µM each of the deoxyribonucleotide triphosphates (dATP, dCTP, dTTP, dGTP) and 0.5 µM of each primer. Samples were overlaid with several drops of paraffin to prevent condensation. In each cycle, samples were heated to 91.5 °C for 1 min to denature the DNA, cooled to 55 °C for 1 min to anneal the primer and heated to 70 °C for 6 min to activate the polymerase. Aliquots of the reaction mix (10 µl) were electrophoresed in a 1% agarose gel, transferred to Gene Screen Plus (Dupont) and hybridized in 50% formamide, 1M NaCl and 1% SDS. The hybridization probe was <sup>32</sup>P-labelled to a specific



activity of 1 × 10<sup>9</sup> c.p.m. µg<sup>-1</sup> using a Multiprime kit (Amersham) and was used at 5 × 10<sup>5</sup> c.p.m. ml<sup>-1</sup>. Filters were washed in 2 × SSC, then 2 × SSC-1% SDS and finally in 0.1 × SSC at 65 °C. Exposure time, 2 (a) and 5 hours (b).

**FIG. 3 a**, Southern analysis of *KpnI* digests of genomic DNA from cloned ES cells hybridized with probe 1. Lane 1, clone showing the 5.4-kb band indicative of a homologous recombination event (for details see text). Lanes 2, 3 and 4, clones in which no mutant *Hox 1.1* allele can be detected. **b**, Southern analysis of an ES-cell clone from pool 1. The clone was grown to a high density to minimize contamination by DNA from feeder cells. Southern blots of *KpnI*-digested DNA, hybridized to probe 1 (lane 1) and 2 (lane 2), show the expected bands at 5.4-kb and 8.6-kb, respectively. **c**, Southern analysis of an ES cell clone from pool 7. Genomic DNA was digested with *StuI/KpnI* (lane 1) or *KpnI* (lane 2) and hybridized to probe 3. **d**, Schematic representation of the mutant and normal *Hox 1.1* allele and the expected restriction fragments, indicating the replacement of the *EcoRI* site by a *KpnI* site in the second exon of the mutant allele. Abbreviations K, *KpnI*; E, *EcoRI*; F, *FspI*; S, *StuI*.

**METHODS.** Single ES cells were picked with glass capillaries<sup>17</sup> and plated into 96-well dishes. ES-cell colonies were subsequently expanded into 1.5-cm, 24-well dishes and then into 6-cm tissue-culture dishes. DNA was isolated from half of these cells and the other half was stored in liquid nitrogen. During the entire cloning procedure ES cells were cultured on feeder cells. Aliquots (10 µg) of genomic DNA were digested with *KpnI*, electrophoresed in 0.8% agarose gels and transferred to Gene Screen Plus. Filters were hybridized to <sup>32</sup>P-labelled probes and washed under the conditions described in Fig. 2.





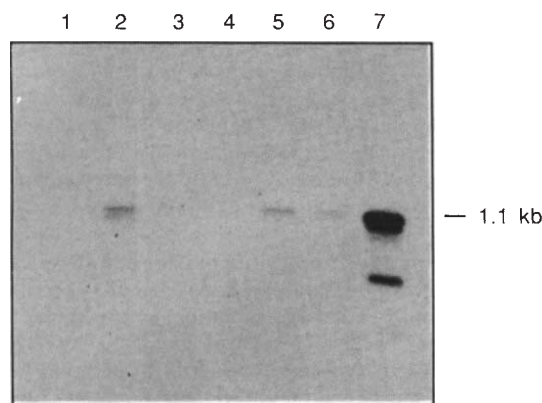


FIG. 4 Tail biopsy of two litters derived from blastocysts, which were microinjected with the clonal mutant ES cell line described in Fig. 3a and b (lanes 1–4, litter lanes 5–7, litter 2). 100 ng of genomic DNA was subjected to 30 PCR cycles as described in Fig. 2. Lanes 2, 5, 6 and 7 exhibit the amplification of the 1.1-kb fragment indicative of the mutant allele (see Figs 1 and 2). METHODS. Mice from two litters were separated after weaning and their tails were cut with new razor blades to prevent contamination. DNA was isolated by standard procedures and analysed by PCR as described in Fig. 2. Exposure time was 5 hours.

ized to probe 3 (see Fig. 3d). The resultant 600-bp and 4.27-kb fragments are derived from the mutated and normal alleles, respectively (see Fig. 3d for details). The mutated and the normal allele are represented in equal amounts in the *StuI/KpnI* and the *KpnI* digest (Fig. 3c, lanes 1 and 2) indicating that the ES cells from pool 7 are clonal. A karyotype analysis of the clone derived from pool 1 revealed that approximately 70% of the cells contained 40 chromosomes (data not shown). Therefore we believe that the lower representation of the mutated allele in this line is due to contamination with non-mutated ES cells and does not reflect a chromosomal instability. We estimate that the ratio of homologous to illegitimate recombination is about 1:30, as about 20% of microinjected cells become stable transformants after nuclear microinjection<sup>13</sup> and homologous recombination occurs in 1 per 150 microinjected cells. This is surprisingly high, relative to the results of comparable studies<sup>11,18,19</sup>. As the parameters that affect the frequency of homologous recombination are not well defined, the explanation for this difference must await further studies.

Our results probably reflect two main differences between our studies and earlier ones. First, the ratio of the heterologous to homologous sequences is very low in the construct used for microinjection. It has been demonstrated that the length of homology is directly proportional to the frequency of gene conversion<sup>11</sup>, whereas the length of the heterology is inversely proportional<sup>11,14</sup>. The other factor favouring high frequency could be the sensitivity of our analysis. We avoided potential problems associated with the requirement for marker gene expression by direct analysis of the DNA target with the highly sensitive PCR method which can potentially detect any gene conversion event. This should make *in vivo* mutagenesis an easy and efficient method for dissecting mammalian developmental pathways. Furthermore, it allows the introduction of more specific mutations, because conditions for the PCR analysis can be selected to detect even single base-pair exchanges<sup>10</sup>. These would be particularly desirable in cases where insertional inactivation of a gene leads to dominant lethal mutations. It should also be possible to generate mouse models for human diseases in which single amino acids are exchanged and to alter the expression pattern of a gene by modifying transcription-factor binding sites. Finally, the method is of potential interest for human somatic gene therapy, because it allows the repair of defective genes without the introduction of other foreign DNA sequences.

The cells containing the mutant allele showed no alterations in their morphology or in their growth properties (data not shown). The ES cells were microinjected into blastocysts of C57BL/6 mice and two litters with a total of four chimaeric animals were obtained. To confirm that the chimaeras contained the mutant allele, tail DNA was analysed by PCR (Fig. 4). In all four chimaeric animals, the expected 1.1-kb fragment was seen after amplification with primers 1 and 2 (Fig. 4, lanes 2, 5, 6 and 7). At present all are healthy and show no abnormalities. These results show that homologous recombination of nonselectable genes can be obtained with high frequency in ES cells which can then be used to produce chimaeric animals. As relatively few cells are required for microinjection, PCR analysis and cloning, the number of individual cell manipulations is small. This reduces the possible accumulation of chromosomal aberrations and maintains a high frequency of germ line contribution. Thus, mouse mutants lacking an intact *Hox 1.1* gene can be bred from chimaeric animals in order to determine gene function. By analogy, functional analysis through mutation of any gene may now be possible. □

Received 17 October 1988; accepted 12 January 1989.

1. Dressler, G. R. & Gruss, P. *Trends Genet.* **4**, 214–219 (1988).
2. Holland, P. W. H. & Hogan, B. L. M. *Genes Dev.* **2**, 773–782 (1988).
3. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1–3 (1988).
4. Balling, R., Deutsch, U. & Gruss, P. *Cell* **55**, 531–535 (1988).
5. Martin, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7634–7638 (1981).
6. Bradley, A., Evans, M., Kaufmann, M. H. & Robertson, E. *Nature* **309**, 255–256 (1984).
7. Doetschmann, T. C., Eistetter, M., Katz, M., Schmidt, W. & Kemler, R. *J. Embryol. exp. Morph.* **87**, 27–45 (1985).
8. Robertson, E., Bradley, A., Kuehn, M. R. & Evans, M. J. *Nature* **323**, 445–448 (1986).
9. Gossler, A., Doetschmann, T., Korn, R., Serfling, E. & Kemler, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9065–9069 (1986).
10. Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
11. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503–512 (1987).
12. Doetschmann, T. *et al. Nature* **330**, 576–578 (1987).
13. Capecchi, M. R. *Cell* **22**, 479–488 (1980).
14. Letson, A. & Liskay, R. M. *Genetics* **117**, 759–769 (1987).
15. Colberg-Poley, A. M., Voss, S. D., Chowdhury, K. & Gruss, P. *Nature* **314**, 713–717 (1985).
16. Kessel, M., Schulze, F., Fibi, M. & Gruss, P. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5306–5310 (1987).
17. Rudnicki, M. A. & McBurney, M. W. in *Teratocarcinomas and Embryonic Stem Cells* (ed. Robertson, E. J.) 19–49 (IRL, Oxford, 1987).
18. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. *Nature* **317**, 230–234 (1985).
19. Thomas, K. R., Folger, K. R. & Capecchi, M. R. *Cell* **44**, 419–428 (1986).

ACKNOWLEDGEMENTS. We thank Dr Rolf Kemler for the D3 cell line and Drs Colin Stewart, Lindsay Williams and Erwin Wagner for the introduction to blastocyst injection. We also thank Sabine Geisendorf for technical assistance, Dr Michael Kessel for helpful discussions and Drs Greg Dressler and Corinne Lobe for reading the manuscript. This work was supported by the Max-Planck-Society and the Deutsche Forschungsgemeinschaft.

## Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells

Alexandra L. Joyner, William C. Skarnes & Janet Rossant

Division of Molecular and Development Biology,  
Mount Sinai Hospital Research Institute, 600 University Avenue,  
Toronto, Ontario, Canada M5G 1X5 and  
Department of Medical Genetics, University of Toronto

A FULL understanding of the function of genes that control developmental events can be obtained only by a combination of molecular and mutational analysis. One putative developmental gene is the mouse engrailed-like gene *En-2*, which was isolated by virtue of its extensive homology to *Drosophila engrailed*, which contributes to the control of segmentation in the developing insect<sup>1</sup>. Our hybridization analysis *in situ* has revealed that expression of *En-2* is restricted to a specific domain of the developing central nervous system from 8 days of development on<sup>2</sup>, indicating a role

for the gene in establishing spatial domains in the brain. Unfortunately no *En-2* mutations are available to elucidate further its function in development. To this end, we report here the isolation of three pluripotent embryonic stem cell lines in which one copy of the homoeobox-containing gene, *En-2*, has been altered by homologous recombination.

Two groups have recently reported homologous recombination of the hypoxanthine-guanine phosphoribosyl transferase (HPRT) gene in embryonic stem (ES) cells<sup>3,4</sup>. Pluripotent ES cells<sup>5,6</sup> can be genetically manipulated in culture and then returned to the embryonic environment<sup>7,8</sup> where they contribute to all tissues of resulting chimaeras, thus allowing the introduction of mutations generated in culture into the mouse gene pool. In the case of the X-linked HPRT gene, it was possible to select directly, in male ES cells, for mutations in the gene<sup>3,4</sup>. To use this strategy to generate mutations in non-selectable autosomal genes, however, some other means of distinguishing homologous integration events must be devised.

Our strategy for introducing mutations to *En-2*, which is similar to one suggested recently<sup>9</sup>, is based on the integration of a *neo*-expressing replacement-type vector, with the use of polymerase chain reaction (PCR)<sup>10</sup> to identify targeted integration events. The vector contains the bacterial neomycin-resistance gene (*neo*) driven by a 500-base pair (bp) fragment of the human  $\beta$ -actin promoter<sup>11</sup> ( $\beta$ -*neo*) flanked on one side by a 3-kilobase (kb) *En-2* fragment and on the other by a 700-bp *En-2* fragment (Fig. 1). Two 500-bp sequences, in the homoeobox-containing exon and the preceding intron, are deleted in this vector and replaced by  $\beta$ -*neo*.

To identify a single targeted colony in a pool of G418-resistant (*neo*<sup>r</sup>) colonies, we took advantage of the PCR technique. Two 20-bp oligonucleotide primers were made, one specific for a 3' *neo* sequence and the other for an opposing *En-2* sequence downstream of the 700-bp *En-2* sequence present in the targeting vector (Fig. 1). Only if the vector integrates correctly through

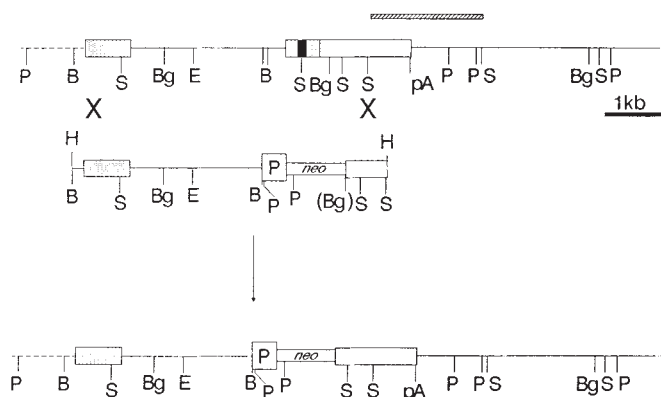


FIG. 1 Alteration of *En-2* by homologous recombination. Schematic diagram showing the *En-2* locus in ES cells (top), the targeting vector (middle) and the predicted sequences present in *En-2* after targeted integration of the vector (bottom). The *En-2* intron and 3' flanking sequences are shown as a line, 5' sequences that have not been fully characterized as a dashed line, exon non-coding sequences as open boxes, exon coding sequences as stippled boxes and the homoeobox as a dark box. The 5' end of the vector joins intron sequences (S. Noble-Topham and A.L.J., unpublished observations). The human  $\beta$ -actin promoter fragment (P) is a 520-bp *Sst*/*Sau*3A fragment from -471 to +49<sup>11</sup> and *neo* is a 1-kb *Bgl*II/*Sma*I fragment which includes the entire coding sequence<sup>18</sup>. To limit the detection of non-homologous integration events, a polyadenylation signal was not provided for the *neo* gene. When the SV40 polyadenylation signal was provided however, we observed a maximal twofold increase in the number of *neo*<sup>r</sup> colonies. The *Hind*III (H) restriction sites at the end of the targeting vector are in the plasmid (pUC) sequences adjacent to the vector. The cross-hatched box above the *En-2* genomic sequences indicates the fragment used in the Southern blot analysis of targeted cell DNA (see Fig. 2). The primers used for the PCR are shown as arrows and the lines between them indicate the fragment amplified by the PCR in the DNA from targeted cells.

homologous recombination should repeated rounds of the PCR with DNA from such cells result in amplification of an 800-bp DNA fragment. To test this, and to optimize our PCR conditions, we introduced into ES cells a vector containing *neo* and *En-2* 3' DNA extending past the *En-2* primer sequence. Using as few as 100 transformed cells, either alone or combined with up to 10<sup>4</sup> negative cells, and 40 rounds of PCR, we detected an 800-bp *En-2* amplified DNA band on gels stained with ethidium bromide.

To avoid the extensive rounds of colony pooling required for standard sib-selection schemes, we devised a procedure that allows rapid identification and isolation of targeted colonies. Male D3 ES cells<sup>12</sup> were electroporated with vector DNA and plated at a dilution to produce about 10 *neo*<sup>r</sup> colonies per dish. G418 was added 48 hours later and cells were maintained under selection for 10–12 days. Portions (100–500 cells) of each colony on a plate were then picked, pooled and subjected to PCR (Table 1) after which aliquots of the reaction mixtures were run on an agarose gel and examined for the expected 800-bp *En-2* fragment. Once a targeted colony was identified within a pool, individual colonies from the original plate were passed and a portion of each (~100 cells) was again subjected to the PCR. By this means, individual targeted colonies were identified and isolated for further culture within 2 weeks of the start of the experiment without the need to pool colonies in culture.

Three separate experiments were performed and targeted

TABLE 1 Homologous recombination into the mouse *En-2* gene

| Experiment | No. D3 cells treated ( $\times 10^7$ ) | No. of <i>neo</i> <sup>r</sup> colonies screened | No. colonies targeted | No. targeted ES clones isolated |
|------------|----------------------------------------|--------------------------------------------------|-----------------------|---------------------------------|
| 1          | 2.5                                    | 500                                              | 2                     | 1                               |
| 2          | 2.5                                    | 400                                              | 1                     | 0                               |
| 3          | 2.5                                    | 400                                              | 2                     | 2                               |

Efficiency of homologous recombination: 1 targeted event per 260 *neo*<sup>r</sup> ES colonies; 1 targeted event per  $1.5 \times 10^7$  treated ES cells.

$2.5 \times 10^7$  D3 ES cells resuspended in 8 ml PBS were electroporated with 200  $\mu$ g *Hind*III digested targeting vector DNA (see Fig. 1) with a Bio Rad Gene Pulser using 500  $\mu$ F and 240 V. The treated cells were seeded onto 40 35-mm tissue culture dishes coated with gelatin in Buffalo rat liver cell-conditioned medium<sup>16</sup>. The medium was replaced 48 h later with medium containing 100  $\mu$ g ml<sup>-1</sup> G418 (GIBCO). The G418-containing medium was subsequently changed every 1–2 days. A concentration of 100  $\mu$ g ml<sup>-1</sup> G418 was chosen because it gave the highest number of *neo*<sup>r</sup> ES cell colonies after electroporation with a  $\beta$ -*neo* vector while still giving no colonies in untreated control cells. The plating efficiency of the treated versus untreated cells was determined by plating ~2,000 cells on a gelatin-coated 35-mm dish in BRL-conditioned medium and fixing, staining and counting the colonies present after 10 days. Typically, 50% of the cells survived electroporation and the plating efficiency of our ES cells under these conditions was 10%.

Preparation of the cells for the PCR was essentially as described<sup>8</sup>. Cells were picked in PBS, spun down in an Eppendorf tube and all but ~5  $\mu$ l of the PBS removed. The cells were resuspended in 50  $\mu$ l water, frozen on dry ice for up to 2 h, and then heated at 94 °C for 8 min. Proteinase K was added (1  $\mu$ g), paraffin oil (40  $\mu$ l) was layered over the mixture, which was then incubated at 50 °C for 1.5 h. Five  $\mu$ l of the 10 $\times$  reaction buffer recommended by Perkin Elmer Cetus was added and the mixture was heated at 95 °C for 6 min, and then incubated at 4 °C until the PCR was started (within 30 min). A cocktail containing the remaining components was then added to each mixture such that the final reaction volume was 50  $\mu$ l and contained 1  $\mu$ g of each primer (*neo*, TCTCATGCTGGAGTCTTCG; *En-2*, ACACATGTCTCTAGATTTCG), 3U *Taq* polymerase (Perkin Elmer Cetus) and 1  $\mu$ l of a mixture of dATP, dCTP, dGTP and dTTP all at 10 mM. PCR was run for 40 cycles using a thermal cycler (Perkin Elmer Cetus) and a denaturation time of 2 min at 94 °C, annealing time of 2 min at 50 °C and extension time of 10 min at 64 °C. The reaction mixture (10  $\mu$ l) was then run on a 1% agarose gel, stained with ethidium bromide, transferred to Gene Screen nylon membrane for 5 h and probed as described previously<sup>17</sup> using the 700-bp *Bgl*II/*Sst*I *En-2* fragment that should be amplified in a targeted colony. Targeted colonies that were successfully isolated and expanded on dishes containing a primary embryonic fibroblast feeder layer were frozen as soon as possible and used for blastocyst injection.



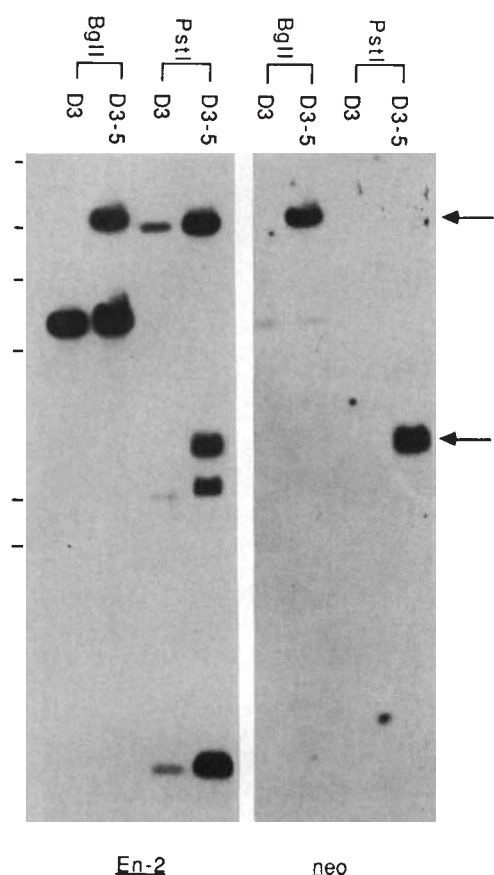


Fig. 2 Southern blot autoradiograms of digested DNA from the parent D3 cell line and one *neo*<sup>r</sup> targeted cell line D3-5, probed with an *En-2* probe (left) and with a *neo* probe (right). The arrows indicate the new bands detected at the targeted locus and the lines on the left indicate the migration positions of *Hind*III-digested  $\lambda$  DNA markers (from top to bottom: 23-, 9.4-, 6.4-, 4.3-, 2.3- and 2.0-kb).

**METHODS.** D3 DNA (5  $\mu$ g) and DNA (10  $\mu$ g) from the targeted colony D3-5 were digested with *Bgl*II or *Pst*I, run on a 0.9% agarose gel, blotted onto Gene Screen and analysed by the Southern blot technique as previously described<sup>17</sup>. The *neo* probe was the 1-kb *neo* fragment present in the targeting vector shown in Fig. 1. The *En-2* probe was the 2-kb *Sst*I fragment 'mp2' probe previously described<sup>17</sup> and shown in Fig. 1.

events were detected by PCR in all three (Table 1). The frequency of targeted integration was about 1 event per 300 *neo*<sup>r</sup> colonies. Three of the five mutant cell lines were recovered and expanded, and each seemed to resemble the founder D3 cell line in morphology, growth rate and karyotype. Targeted integration into the *En-2* locus was confirmed by Southern blot analysis of DNA from each cell line. DNA from the parent D3 cell line and from each of the three targeted lines was digested with the restriction enzymes *Bgl*II, *Pst*I, *Eco*RI and analysed with an *En-2* probe and a *neo* probe. DNA from each of the targeted lines produced the predicted restriction patterns for a homologous recombination event in one allele of *En-2* (see Fig. 2). Thus, the targeting vector had integrated by means of homologous recombination in the predicted manner with no apparent major rearrangements of either vector or host DNA.

The approach described here should be applicable to any gene for which the frequency of homologous recombination is at least 1 per 1,000 of the *neo*<sup>r</sup> colonies obtained. In the case of *En-2*, our frequency of 1 per 260 *neo*<sup>r</sup> colonies is higher than that obtained for HPRT using a similar *neo*-expressing replacement-type vector<sup>4</sup>, which was at best 1 targeting event per 1,000 *neo*<sup>r</sup> colonies. This apparent enrichment for homologous recombination events may be due to the fact that our *neo* construct does not include a polyadenylation signal (Fig. 1) and contains a weaker promoter than that used for HPRT (W.C.S. and A.L.J., unpublished observations). It is not yet clear what

TABLE 2 Chimaera production from *En-2*-targeted ES cell lines

| Blastocyst strain | Cell line | No. blastocysts injected | No. progeny (% injected) | No. chimaeras (% progeny) |
|-------------------|-----------|--------------------------|--------------------------|---------------------------|
| CD1               | D3        | 60                       | 41 (66)                  | 23 (56)                   |
|                   | D3-5      | 62                       | 41 (66)                  | 5 (12)                    |
|                   | D3-15     | 46                       | 33 (71)                  | 7 (21)                    |
|                   | D3-39     | 93                       | 47 (50)                  | 13 (27)                   |
| C57BL/6J          | D3-5      | 55                       | 14*(25)                  | 2†                        |
|                   | D3-15     | 26                       | 4 (15)                   | 0‡                        |
|                   | D3-39     | 23                       | 6§(26)                   | 4                         |

Groups of 10–20 ES cells from each cell line were injected into blastocysts from either outbred CD1 (Charles River, Quebec) or inbred C57BL/6J (Jackson Lab.) mice. Injected blastocysts were transferred to the uteri of pseudo-pregnant recipients and allowed to proceed to term. Chimaerism in resulting offspring was assessed by coat pigmentation. The D3 line, derived from 129J mice<sup>12</sup>, gives an agouti coat colour in either an albino (CD1) or black, non-agouti background (C57BL/6J).

\* 7 of these represent fetuses dissected at 12.5 days of development.

† These two chimaeras were identified by PCR in the group of 7 12.5-day fetuses. One was noticeably haemorrhagic and would probably not survive to term.

‡ Three offspring died at birth and were eaten, precluding analysis for mosaicism.

§ One offspring died 5 days after birth, precluding analysis for mosaicism.

factors influence the apparent frequency of homologous recombination at different loci. One factor may be the level of expression of the targeted genes in ES cells, particularly for strategies that require expression of *neo*. The steady-state level of *En-2* mRNA is however, very low in ES cells (A.L.J., unpublished observations) and in PSA-I teratocarcinoma cells<sup>1</sup>. For loci that have a very low frequency of homologous recombination, it will be necessary to incorporate additional means of enriching for homologous recombination events, such as the three methods recently described<sup>13–15</sup>.

Using ES cell lines with one mutant *En-2* allele, it should be possible to introduce the mutation into the mouse germ line and analyse its effect on embryonic development. All three mutant cell lines were injected into both outbred CD1 and inbred C57BL/6J blastocysts (Table 2). The four male CD1 chimaeras from D3-5 were test bred, but we found no progeny derived from the ES cells in more than 50 offspring from each male. This was not entirely unexpected because the contribution of D3-5 to somatic tissues was very low. So far we have obtained only a few live-born young from the C57BL/6J series, although three had a large ES cell contribution in the coat. These mice are still too young to test for germ-line mosaicism. More chimaera data or germ-line transmission of the mutant *En-2* allele are necessary to determine whether the low number of chimaeras obtained is due to a dominant effect of the mutation on development.

The use of PCR and our pooling procedure should be applicable to the final screening stage of any homologous recombination scheme and help in the production of many mouse germ-line mutations. □

Received 12 December 1988; accepted 17 January 1989.

1. Joyner, A. L. & Martin, G. R. *Genes Dev.* **1**, 29–38 (1987).
2. Davis, C. A., Noble-Topham, S. E., Rossant, J. & Joyner, A. L. *Genes Dev.* **2**, 361–371 (1988).
3. Doetschman, T. et al. *Nature* **330**, 576–578 (1987).
4. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503–512 (1987).
5. Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154–156 (1981).
6. Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7634–7638 (1981).
7. Gossler, A., Doetschman, T., Korn, A., Serfling, E. & Kemler, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9065–9069 (1986).
8. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. *Nature* **323**, 445–448 (1986).
9. Kim, H.-S. & Smithies, O. *Nucleic Acids Res.* **16**, 8887–8903 (1988).
10. Sakai, R. K., Scharf, S., Faloona, F., Mulliswork, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. *Science* **230**, 1350–1354 (1985).
11. Frederickson, R. M., Micheau, M. R., Iwamoto, A. & Miyamoto, N. *Nucleic Acids Res.* **17**, 253–270 (1989).



12. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. *J. Embryol. exp. Morph.* **87**, 27–45 (1985).
13. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348–352 (1988).
14. Jasini, M. & Berg, P. *Genes Dev.* **2**, 1353–1363 (1988).
15. Sedivy, J. M. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
16. Smith, A. G. & Hooper, M. L. *Dev. Biol.* **121**, 1–9 (1987).
17. Joyner, A. L., Kornberg, T., Coleman, K. G., Cox, D. & Martin, G. R. *Cell* **43**, 29–37 (1985).
18. Southern, P. J. & Berg, P. *Molec. appl. Genet.* **1**, 327–341 (1982).

ACKNOWLEDGEMENTS. We thank B. A. Auerbach for her technical assistance with the ES cells and D. Nallathannan for help with molecular procedures. We also thank A. Gossler, O. Smithies and T. Doetschman for discussions, R. Kemler for the D3 cell line and N. Miyamoto for the  $\beta$ -actin promoter fragment. This work was supported by the National Cancer Institute and the Medical Research Council of Canada.

## Preferential mutation of paternally derived RB gene as the initial event in sporadic osteosarcoma

Junya Toguchida<sup>\*†</sup>, Kanji Ishizaki<sup>\*</sup>, Masao S. Sasaki<sup>\*†</sup>, Yusuke Nakamura<sup>§</sup>, Mituo Ikenaga<sup>\*</sup>, Mitsuo Kato<sup>\*</sup>, Masayuki Sugimoto<sup>\*†</sup>, Yoshihiko Kotoura<sup>†</sup> & Takao Yamamuro<sup>†</sup>

<sup>\*</sup>Radiation Biology Center and <sup>†</sup>Department of Orthopaedic Surgery, Faculty of Medicine, Kyoto University, Sakyo-ku, Kyoto 606, Japan  
<sup>§</sup>Howard Hughes Medical Institute and Department of Human Genetics, University of Utah Health Sciences Center, 603 Wintrobe Building, Salt Lake City, Utah 84132, USA

SUCCESSIVE loss of function of both alleles of the retinoblastoma susceptibility gene (RB) on human chromosome 13 seems to be critical in the development of retinoblastoma<sup>1–3</sup> and osteosarcoma<sup>4–6</sup>. In cases where the tumour is familial and susceptibility is inherited, a mutation in one of the alleles is carried in the germline<sup>7</sup>. We have recently shown that cytogenetically visible germline mutations are usually in the paternally derived gene<sup>8</sup>. Such a bias would not be expected for sporadic (non-familial) tumours, where both mutations occur in somatic tissue, but there has been some indication of a bias towards initial somatic mutation in the paternally derived gene on chromosome 11 in sporadic Wilms tumour<sup>9</sup>. We have now examined 13 sporadic osteosarcomas and find evidence which indicates that in 12 cases the initial mutation was in the paternal gene, suggesting the involvement of germinal imprinting in producing the differential susceptibility of the two genes to mutation.

Our approach to the determination of the parental origin of the RB mutation stems from the two-mutation theory<sup>10,11</sup>, in which the tumour is thought to develop as a result of the unmasking of the initial mutation by a subsequent mutation that eliminates the wild-type allele. In many cases this process can be recognized as a loss of heterozygosity at polymorphic loci around the RB locus<sup>12</sup>, which enables the initial mutation to be identified. We have chosen sporadic osteosarcoma, in which the initial mutation is supposed to occur in somatic cells, to establish the parental origin of the chromosome harbouring the initial mutation in the RB gene, using polymorphic DNA probes to chromosome 13 and complementary DNA probes derived from the RB gene. DNA polymorphisms represent normal variations in genomic DNA sequences and can be used to distinguish the maternal and paternal copies of a chromosome. Comparison of DNA from constitutional and tumour tissue can reveal which copy carries the wild-type allele and has been lost during tumorigenesis.

A total of 13 cases were used to identify the parental origin of the lost chromosome or chromosome segment (Table 1). Tumours were classified into three groups according to structural

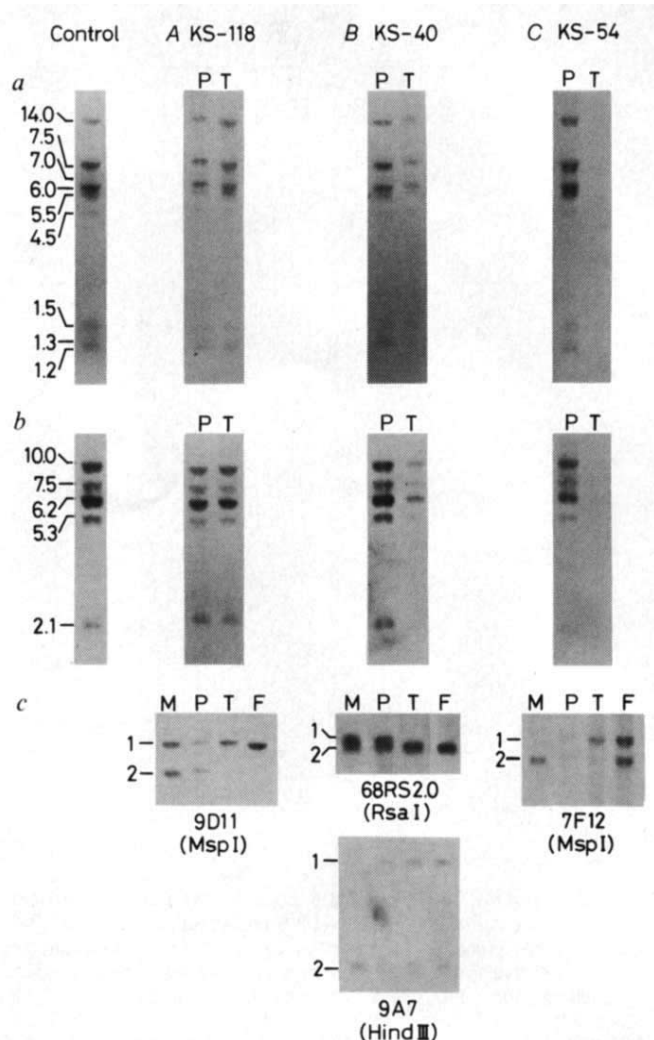


FIG. 1 Autoradiographs of structural analysis of the RB gene and restriction fragment-length polymorphism analysis of chromosome 13. *a* and *b*, Southern analysis of *Hind*III-digested genomic DNA using cDNA probes of the RB gene, RB-1(*a*) and RB-5(*b*). Numbers on the left indicate the length of each fragment in kilobases. *c*, Southern analysis of polymorphic loci on chromosome 13. The polymorphic probes and restriction enzymes used are indicated underneath each gel. Details of these probes are given in the legend to Table 1. Alleles at each locus are designated as 1 or 2 according to size. KS-118, KS-40 and KS-54 are representative examples from tumour groups A, B and C respectively, as described in the text.

METHODS. High-molecular weight DNA was isolated from peripheral leukocytes of patients (P), their mothers (M), their fathers (F) and from their osteosarcoma tumour tissue (T); DNA from the peripheral leukocytes of a healthy donor was used as a control. Restriction endonuclease digestion of these samples followed by analysis on Southern blots has been described<sup>29</sup>. *a* and *b*, The cDNA probes of the RB gene<sup>2</sup>, RB-1 and RB-5 (gift from Dr W.-H. Lee), span 1.6 kilobases at the 5' end and 3.5 kilobases at the 3' end of the RB gene respectively, with a 0.4-kilobase overlap between them. They hybridize with 9 and 5 *Hind*III-digested genomic DNA fragments respectively, and share a common 7.5-kilobase fragment.

abnormalities shown in the RB gene. Figure 1 shows examples typical of each group. Group A tumours are from 6 osteosarcoma cases (Table 1) in which no structural abnormality of the RB gene was detected in either constitutional or tumour tissues. Loss of heterozygosity was found in tumours for all informative polymorphic loci, the missing sequences being lost from the maternally derived chromosome in five cases and from the paternally derived chromosome in one case. Group B consists of 3 tumours (Table 1) in which constitutional heterozygosity has been lost within a region that includes the RB locus, but has been retained outside this region. No structural abnormality was detected in the remainder of the RB gene. The parental

<sup>†</sup>To whom correspondence should be addressed.

TABLE 1 Alleles present at polymorphic loci on chromosome 13 in osteosarcoma patients and their parents

| Case number | DNA    | Probe |         |         |      |     |     |       | Parental origin of lost allele |
|-------------|--------|-------|---------|---------|------|-----|-----|-------|--------------------------------|
|             |        | 7F12  | 123M1.8 | 68RS2.0 | 9D11 | 1E8 | 9A7 | MHZ47 |                                |
| A           | KS-32  | M     |         |         |      | 1,2 | 1,2 |       |                                |
|             |        | P     | CH      | ND      | CH   | 1,2 | 1,2 | ND    |                                |
|             |        | T     |         |         |      | 1   | 1   |       | F                              |
|             |        | F     |         |         |      | 1,2 | 2   |       |                                |
|             | KS-78  | M     | 1,2     |         |      | 1   | 2   |       |                                |
|             |        | P     | 1,2     | ND      | ND   | 1,2 | 1,2 | ND    |                                |
|             |        | T     | 1       |         | CH   | 2   | 1   |       | M                              |
|             |        | F     | 1,2     |         |      | 1,2 | 1,2 |       |                                |
|             | KS-118 | M     |         |         | 1,2  | 1,2 |     | 2     |                                |
|             |        | P     | CH      | ND      | 1,2  | 1,2 | CH  | 1,2   |                                |
|             |        | T     |         |         | 1    | 2   |     | 1     | M                              |
|             |        | F     |         |         | 1    | 1,2 |     | 1     |                                |
| B           | KS-140 | M     |         |         |      |     |     | 1,2   |                                |
|             |        | P     | CH      | CH      | ND   | CH  | CH  | 1,2   |                                |
|             |        | T     |         |         | CH   |     |     | 2     | M                              |
|             |        | F     |         |         |      |     |     | 2     |                                |
|             | KS-147 | M     |         | 1       | 1    |     |     |       |                                |
|             |        | P     | CH      | 1,2     | 1,2  | CH  | CH  | CH    |                                |
|             |        | T     |         | 2       | 2    |     |     |       | M                              |
|             |        | F     |         | 1,2     | 1,2  |     |     |       |                                |
|             | KS-159 | M     | 1,2     |         | 1,2  |     |     |       |                                |
|             |        | P     | 1,2     | CH      | 1,2  | CH  | CH  | CH    |                                |
|             |        | T     | 2       |         | 2    |     |     |       | M                              |
|             |        | F     | 1,2     |         | 1,2  |     |     |       |                                |
| C           | KS-40  | M     |         | 1,2     | 1,2  |     | 2   |       |                                |
|             |        | P     | CH      | 1,2     | 1,2  | CH  | 1,2 | ND    |                                |
|             |        | T     |         | 2       | 2    |     | 1,2 |       | M                              |
|             |        | F     |         | 2       | 2    |     | 1,2 |       |                                |
|             | KS-66  | M     |         | 1       | 1    | 1,2 | 2   |       |                                |
|             |        | P     | CH      | 1,2     | 1,2  | 1,2 | 1,2 | ND    |                                |
|             |        | T     |         | 2       | 2    | 2   | 1,2 |       | M                              |
|             |        | F     |         | 1,2     | 2    | 1,2 | 1,2 |       |                                |
|             | KS-112 | M     | 2       |         |      | 1,2 | 2   |       |                                |
|             |        | P     | 1,2     | CH      | CH   | 1,2 | 1,2 | ND    |                                |
|             |        | T     | 1       |         | CH   | 1,2 | 1,2 |       | M                              |
|             |        | F     | 1,2     |         |      | 1,2 | 1   |       |                                |
| C           | KS-22  | M     |         |         | 1,2  |     |     |       |                                |
|             |        | P     | CH      | CH      | 1,2  | CH  | CH  | CH    |                                |
|             |        | T     |         |         | 1    |     |     |       | M                              |
|             |        | F     |         |         | 1    |     |     |       |                                |
|             | KS-30  | M     |         |         | 1,2  | 1,2 |     |       |                                |
|             |        | P     | CH      | CH      | 1,2  | 1,2 | CH  | ND    |                                |
|             |        | T     |         |         | 1    | 2   |     |       | M                              |
|             |        | F     |         |         | 1    | 1,2 |     |       |                                |
|             | KS-54  | M     | 2       |         |      | 1,2 |     |       |                                |
|             |        | P     | 1,2     | ND      | ND   | 1,2 | CH  | ND    |                                |
|             |        | T     | 1       |         | CH   | 1   |     |       | M                              |
|             |        | F     | 1,2     |         |      | 1,2 |     |       |                                |
| C           | KS-81  | M     |         | 2       |      | 1,2 |     | 2     |                                |
|             |        | P     | CH      | 1,2     | CH   | 1,2 | CH  | 1,2   |                                |
|             |        | T     |         | 1       |      | 1   |     | 1     | M                              |
|             |        | F     |         | 1,2     |      | 1,2 |     | 1     |                                |

Alleles of polymorphic loci were determined by Southern hybridization as described in the legend to Fig. 1. A, B and C indicate tumour groups as described in the text. Designations for DNA samples are as described in the legend to Fig. 1. All tumour tissues were derived from primary lesions, except KS-40 and KS-81 (from lung metastasis). The designation and location of sequences homologous with chromosome 13 in p7F12, p9D11, p1E8 and p9A7 are described in ref. 27. p123M1.8 and p68RS2.0 are polymorphic probes within the RB locus<sup>28</sup>, and pMHZ47 recognizes the homologous sequence located near the telomere of chromosome 13 (Y. N., unpublished data). Probes 7F12, 1E8, 9D11, 9A7, 123M1.8 revealed two allele polymorphisms and the longer allele was designated 1 and the shorter designated 2. Probes 68RS2.0 and MHZ47 revealed a variable number of tandem repeats, and several different-length alleles were observed in normals (our unpublished data); designation of each allele of these two probes was 1 and 2 according to length. The order of these markers on chromosome 13q is cen-7F12-(123M1.8-68RS2.0)-9D11-1E8-9A7-(MHZ47)-qter. The parental origin of the lost allele was determined from the results for the boxed locus: F, father; M, mother. CH, constitutional homozygosity; ND, not determined.

origin of the deleted region was maternal in all cases. Group C comprises four tumours (Table 1): no structural abnormality was found in constitutional cells, but in the tumours there were homozygous deletions of the RB gene and loss of heterozygosity at all informative polymorphic loci, with the paternally derived allele being retained.

Altogether, in 12 of 13 tumours, the lost chromosome or chromosome segment was maternally derived. This indicates that the first mutation must have occurred preferentially on the paternally derived chromosome 13; the difference compared with an equal involvement of both chromosomes was highly significant ( $\chi^2 = 9.3$ ,  $P = 0.0023$ ). In group B tumours, the large interstitial deletion involving the RB locus could be the first event, which is followed by inactivation of the other allele by point mutation. When we excluded these tumours because of this uncertainty, the ratio of involvement of maternal to paternal chromosomes was 1:9, still showing a strong bias towards the paternally derived chromosome ( $\chi^2 = 4.9$ ,  $P = 0.027$ ). All tumours were sporadic and solitary, and none of the patients had a family history of eye tumours. Moreover, none of the tumours, including those showing homozygous deletion of the RB gene, had any detectable structural abnormality in their constitutional cells. It is therefore highly probable that the initial mutation was somatic in all cases. But we cannot exclude the possibility that some patients, particularly those in groups A and B, could have inherited a germinal mutation that was not detectable by Southern analysis, and suffered from osteosarcoma without manifesting retinoblastoma, either by chance or as a result of an intrinsic tendency to osteosarcoma.

Preferential involvement of a paternal chromosome in a germinal mutation could be due to the high rate of DNA replication in spermatogenesis<sup>13</sup>. But this may not hold for a somatic mutation, and another causal factor may be involved in the phenomenon we observe here. Moreover, a similar parental origin for germinal and somatic mutations indicates that the difference observed in the germinal mutation might not simply be attributable to a difference in the number of cell divisions in two types of germline cells. If this promotes the bias to paternal origin, then the age of the father would be a risk factor for the new mutation<sup>13</sup>. Such a paternal-age effect was not observed, however, for the paternally derived new mutations in retinoblastoma patients (our unpublished data). The non-random loss of maternal alleles on chromosome 11p in Wilms tumours has been attributed to a transforming gene, which is controlled by the Wilms regulatory gene, is located on the same region of chromosome 11p and whose expression differs according to its parental origin as a result of genomic imprinting<sup>14</sup>. If this is true, its expression in familial cases would depend on the parental origin of the mutation. In familial retinoblastoma, however, the expression of the RB mutation in offspring was the same when the mutation was transmitted from either the father or the mother<sup>15</sup>.

There is a growing body of data indicating a difference in behaviour of maternally and paternally derived autosomal genes<sup>16,17</sup>. Germinal imprinting may be mediated by some epigenetic process such as *de novo* DNA methylation and be carried over to post-zygotic stages<sup>18-22</sup>. The differential modification of genes by germinal imprinting and its possible association with a differential susceptibility to mutation of the RB gene is particularly interesting when considering the preferential involvement of a paternally derived gene in mutagenesis. Such differential susceptibility could be a characteristic of post-meiotic cells as the chromosomes of human sperm are highly vulnerable<sup>23,24</sup>; also a high frequency of mutational mosaicism has been noted in retinoblastoma patients<sup>25</sup>. Loss of heterozygosity of a specific chromosome pair has been reported in many tumours, and the loss of the function controlled by this region has been assumed to be responsible for tumorigenesis<sup>26</sup>. Knowing the parental origin of these lost genes could be important in elucidating the mutational origin of human cancers. □

Received 12 December 1988; accepted 31 January 1989.

1. Friend, S. H. *et al.* *Nature* **323**, 643-646 (1986).
2. Lee, W.-H. *et al.* *Science* **235**, 1394-1399 (1987).
3. Fung, Y.-K. T. *et al.* *Science* **236**, 1657-1661 (1987).
4. Hansen, M. F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 6216-6220 (1985).
5. Dryja, T. P. *et al.* *Am. J. Hum. Genet.* **38**, 59-66 (1986).
6. Toguchida, J. *et al.* *Cancer Res.* **48**, 3939-3943 (1988).
7. Vogel, F. *Hum. Genet.* **52**, 1-54 (1979).
8. Ejima, Y., Sasaki, M. S., Kaneko, A. & Tanooka, H. *Hum. Genet.* **79**, 118-123 (1988).
9. Schroeder, W. T. *et al.* *Am. J. Hum. Genet.* **40**, 413-420 (1987).
10. Knudson, A. G. Jr *Proc. natn. Acad. Sci. U.S.A.* **68**, 820-823 (1971).
11. Cavenee, W. K. *et al.* *Nature* **305**, 779-784 (1983).
12. Cavenee, W. K., Koufos, A. & Hansen, M. F. *Mutat. Res.* **168**, 3-14 (1986).
13. Vogel, F. & Rathenberg, R. in *Advances in Human Genetics* Vol. 5 (eds Harris, H. *et al.*) 223-318 (Plenum, New York, 1985).
14. Wilkins, R. J. *Lancet* **i**, 329-331 (1988).
15. Matsunaga, E. *Hum. Genet.* **62**, 124-128 (1982).
16. Surani, M. A. H., Barton, S. C. & Norris, M. L. *Nature* **308**, 548-550 (1984).
17. Cattaniach, B. M. & Kirk, M. *et al.* *Nature* **315**, 496-498 (1985).
18. Groundie, M. & Conkin, K. F. *et al.* *Science* **228**, 1061-1068 (1985).
19. Hadchouel, M., Farza, H., Simon, D., Tiollais, P. & Pourcel, C. *Nature* **329**, 454-456 (1987).
20. Reik, W., Collocic, A., Norris, M. L., Barton, S. C. & Surani, M. A. *Nature* **328**, 248-251 (1987).
21. Sapienza, C. *et al.* *Nature* **328**, 251-254 (1987).
22. Swain, J. L., Stewart, T. A. & Leder, P. *Cell* **50**, 719-727 (1987).
23. Kamiguchi, Y. & Mikano, K. *Am. J. Hum. Genet.* **38**, 724-740 (1986).
24. Martin, R. H. *et al.* *Hum. Genet.* **77**, 108-114 (1987).
25. Motegi, T. *Hum. Genet.* **61**, 95-97 (1982).
26. Green, A. R. *Br. J. Cancer* **58**, 115-121 (1988).
27. Cavenee, W. K. *et al.* *Am. J. Hum. Genet.* **36**, 10-24 (1984).
28. Wiggs, J. *et al.* *New Engl. J. Med.* **318**, 151-157 (1988).
29. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

ACKNOWLEDGEMENTS. We are most grateful to the patients and their parents who cooperated in this investigation. We also thank Dr W.-H. Lee for providing cDNA probes for the RB gene, Drs T. P. Dryja and W. K. Cavenee for the polymorphic probes, and Drs N. Takada, S. Tatezaki, T. Umeda, K. Ohba, H. Urushidani and M. Shima for providing samples. This work was supported by Cancer Research Grants from the Ministry of Education, Science and Culture, Japan.

## Role for carbohydrate structures in TGF- $\beta$ 1 latency

Kohei Miyazono\* & Carl-Henrik Heldin†

Ludwig Institute for Cancer Research, Box 595, Biomedical Center, S-751 23 Uppsala, Sweden

**TRANSFORMING growth factor- $\beta$  (TGF- $\beta$ )** (reviewed in refs 1-3) is a family of molecules that are made up as disulphide-bonded dimers of at least three different types of homologous polypeptides<sup>4-8</sup>. The active molecules are cleaved from the C termini of precursors<sup>4,6-8</sup>. TGF- $\beta$ 1, like other forms of TGF- $\beta$ , is synthesized and secreted in a latent high relative molecular mass form (L-TGF- $\beta$ 1) from which active TGF- $\beta$ 1 can be released by transient and probably unphysiological acidification<sup>9-12</sup>. The latent complex from human platelets contains one dimeric TGF- $\beta$ 1 molecule, which is noncovalently associated with a disulphide-bonded complex of one dimeric remnant of the precursor and a single molecule of the so-called TGF- $\beta$ 1 binding protein (TGF- $\beta$ 1-BP)<sup>11,13</sup>. We report here that enzymatic removal *in vitro* of the carbohydrate structures in the remnant of the TGF- $\beta$ 1 precursor produces biologically active TGF- $\beta$ 1 from the latent complex, suggesting that carbohydrate structures are of importance in rendering TGF- $\beta$ 1 inactive in the complex *in vivo*.

We followed the activation of L-TGF- $\beta$ 1 with assays for inhibition of [<sup>3</sup>H]thymidine incorporation into porcine aortic endothelial cells<sup>11</sup> and by competition assays with [<sup>125</sup>I]-labelled TGF- $\beta$ 1 binding to receptors<sup>14,15</sup> on normal rat kidney fibroblasts (NRK-49F). Native L-TGF- $\beta$ 1, purified to homogeneity from human platelets<sup>11</sup>, inhibited cell growth at high concentrations, with a half-maximal effect at 1 nM; after a transient exposure to 1 M HCl, the inhibitory activity increased 200-fold (Fig. 1a). Similarly, 20 nM of native L-TGF- $\beta$ 1, but only 0.1 nM of acidified L-TGF- $\beta$ 1, reduced [<sup>125</sup>I]-labelled TGF- $\beta$ 1 binding to NRK cells by 50% (Fig. 1b). The activation was complete,

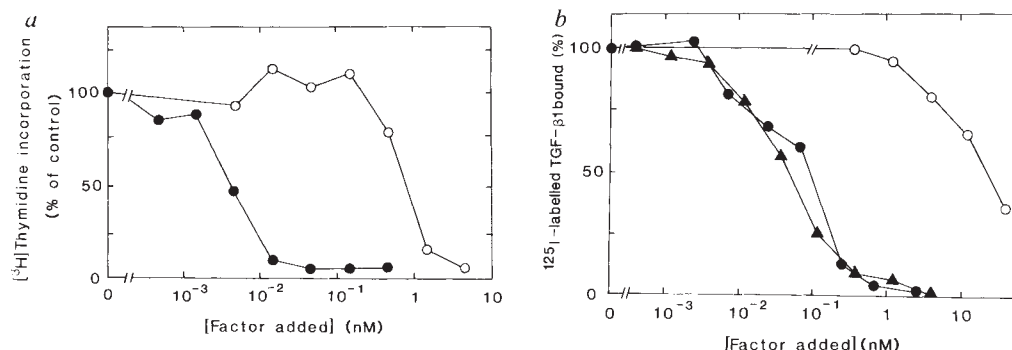
\* Present address: Third Department of Internal Medicine, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan.

† To whom correspondence should be addressed.



FIG. 1 Activation of L-TGF- $\beta$ 1 by acidification. *a*, Inhibition of [ $^3$ H]thymidine incorporation into porcine aortic endothelial cells by native ( $\circ$ ) and transiently acidified ( $\bullet$ ) L-TGF- $\beta$ 1. *b*, Ability of TGF- $\beta$ 1 ( $\blacktriangle$ ), native L-TGF- $\beta$ 1 ( $\circ$ ) and transiently acidified L-TGF- $\beta$ 1 ( $\bullet$ ) to compete with [ $^{125}$ I]-labelled TGF- $\beta$ 1 for binding to NRK cells.

METHODS. The assay for growth inhibition has been described<sup>11</sup>. In the radio-receptor assay for TGF- $\beta$ , NRK-49F cells were grown to confluence in Linbro 12-well plates, and rinsed once with 1 ml of binding buffer (phosphate buffered saline containing 1 mg ml<sup>-1</sup> of bovine serum albumin, 0.9 mM CaCl<sub>2</sub> and 0.49 mM MgCl<sub>2</sub>). Cell cultures were incubated for 2 h at 0 °C in 0.5 ml of binding buffer containing 1–2 ng ml<sup>-1</sup> of [ $^{125}$ I]-labelled TGF- $\beta$ 1 and various concentrations of test samples. Cultures were washed five times in ice-cold binding buffer and cell-associated [ $^{125}$ I]-label extracted by 0.5 ml of 1% Triton X-100, 10% (v/v) glycerol, 20 mM HEPES, pH 7.4, and quantitated by a



$\gamma$ -counter. Radioiodination of TGF- $\beta$ 1 was performed by the lactoperoxidase method<sup>14,15</sup> or by the chloramine T method<sup>25</sup>. Pure L-TGF- $\beta$ 1 was obtained as previously described<sup>11</sup>, and desalted by chromatography on a Superose 6 column. For the acidification of L-TGF- $\beta$ 1, samples were incubated in 1 M HCl for 30 min at 22 °C and then neutralized with 1 M NaOH, as described<sup>11</sup>.

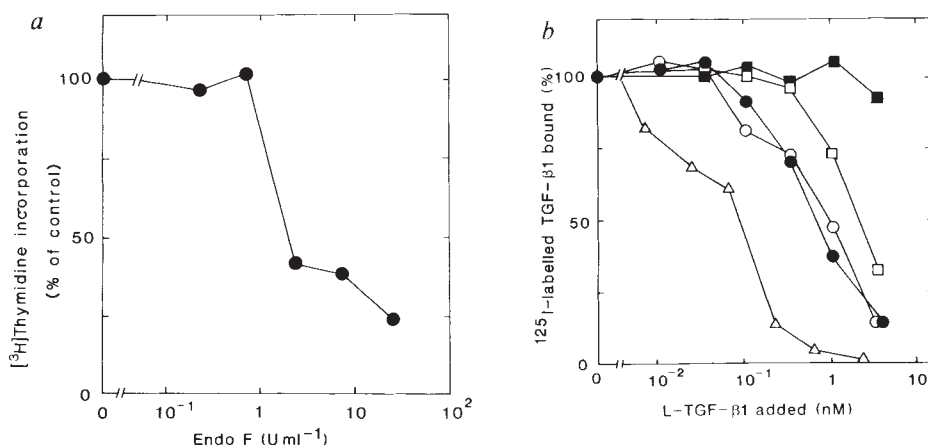
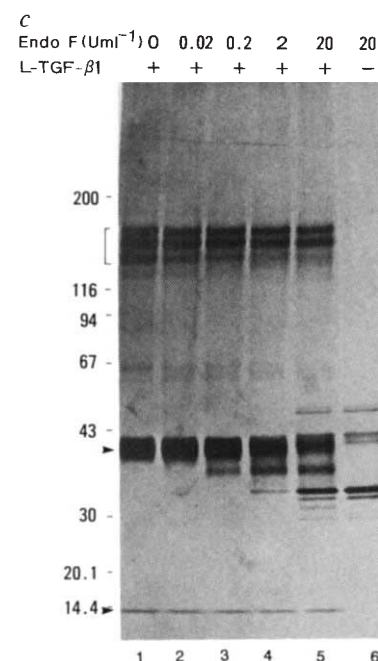


FIG. 2 Activation of L-TGF- $\beta$ 1 by endo F. *a*, L-TGF- $\beta$ 1 digested with endo F at various concentrations was added to the endothelial cells at a concentration of 0.2 nM and assayed for [ $^3$ H]thymidine incorporation into DNA. *b*, Competition binding assays of L-TGF- $\beta$ 1 treated with 0 U ml<sup>-1</sup> ( $\blacksquare$ ), 4 U ml<sup>-1</sup> ( $\square$ ), 10 U ml<sup>-1</sup> ( $\bullet$ ) or 25 U ml<sup>-1</sup> ( $\circ$ ) of endo F with [ $^{125}$ I]-labelled TGF- $\beta$ 1. At these concentrations, endo F has no significant effect in these assays (data not shown). Competition of transiently acidified L-TGF- $\beta$ 1 ( $\triangle$ ) with [ $^{125}$ I]-labelled TGF- $\beta$ 1 is also shown. *c*, L-TGF- $\beta$ 1 digested with endo F was reduced by 50 mM dithiothreitol for 3 min at 100 °C and analysed by SDS-gel electrophoresis and silver staining<sup>26,27</sup>. Lanes 1–5, L-TGF- $\beta$ 1 treated with 0–20 U ml<sup>-1</sup> of endo F; lane 6, the same concentration of endo F as in lane 5, but with no added L-TGF- $\beta$ 1. Bracket on the left, TGF- $\beta$ 1-BP. Arrowheads on the left, positions of TGF- $\beta$ 1 and the untreated N-terminal remnant of the precursor. Arrowheads on the right, degradation forms of the TGF- $\beta$ 1 precursor

as the competition curve of acidified L-TGF- $\beta$ 1 was essentially identical to that of purified TGF- $\beta$ 1 (Fig. 1b).

Because TGF- $\beta$ 1 is produced by most cell types and affects the growth and differentiation of many different cell types<sup>1–3</sup>, its activation must be carefully regulated. Acidification is probably not a physiological mechanism, as a pH of less than 4 is required<sup>11,12</sup>. We investigated the role of carbohydrate structures, as although TGF- $\beta$ 1 does not contain carbohydrate, TGF- $\beta$ 1-BP does<sup>11</sup>, and the precursor remnant has three potential *N*-glycosylation sites<sup>4</sup>, all of which seem to be used<sup>11,16</sup>. Treatment of L-TGF- $\beta$ 1 with endoglycosidase F (endo F), which removes complex and oligomannose (high-mannose) *N*-linked carbohydrate complexes<sup>17,18</sup> led to a dose-dependent activation of L-TGF- $\beta$ 1 (Fig. 2a, b). The concentration (0.5 nM) of L-TGF- $\beta$ 1 pretreated with 10 or 25 U ml<sup>-1</sup> endo F that was needed for a 50% decrease in [ $^{125}$ I]-labelled TGF- $\beta$ 1 binding to NRK cells was five times that needed for L-TGF- $\beta$ 1 completely activated by transient acidification, but 40-fold lower than for native



remnant. Relative molecular masses are shown on the left, in thousands. METHODS. L-TGF- $\beta$ 1 was incubated with endo F from *Flavobacterium meningosepticum* (Boehringer) at 37 °C for 16 h in 50 mM EDTA and 100 mM sodium phosphate, pH 7.4. Incubations were carried out in the absence of detergents and chaotropic agents because these agents activate L-TGF- $\beta$ 1.

L-TGF- $\beta$ 1. Treatment of the complex with endo F causes degradation of the precursor remnant from a relative molecular mass of 40,000 (40K) to 31K (Fig. 2c). The 31K degradation product is probably completely deglycosylated as this is the size obtained after chemical deglycosylation<sup>11</sup> and as it is close to that predicted from the cDNA sequence<sup>4</sup>. This is also consistent with results obtained using recombinant TGF- $\beta$ 1<sup>16</sup>. The 13K TGF- $\beta$ 1 band was, as expected, unaffected by the endo F treatment, as were the 125–160K TGF- $\beta$ 1-BP bands. Endoglycosidase H, which acts on protein-linked oligomannose type oligosaccharides, did not activate L-TGF- $\beta$ 1 (data not shown).

These results indicate that carbohydrate structures in the precursor remnant are involved in binding TGF- $\beta$ 1 in a way in which it remains inactive, and that the removal of these carbohydrate structures leads to the release of active TGF- $\beta$ 1. To investigate whether small carbohydrates could compete with these structures, several monosaccharides were assayed for their ability to activate L-TGF- $\beta$ 1. In the growth inhibition assay, sialic

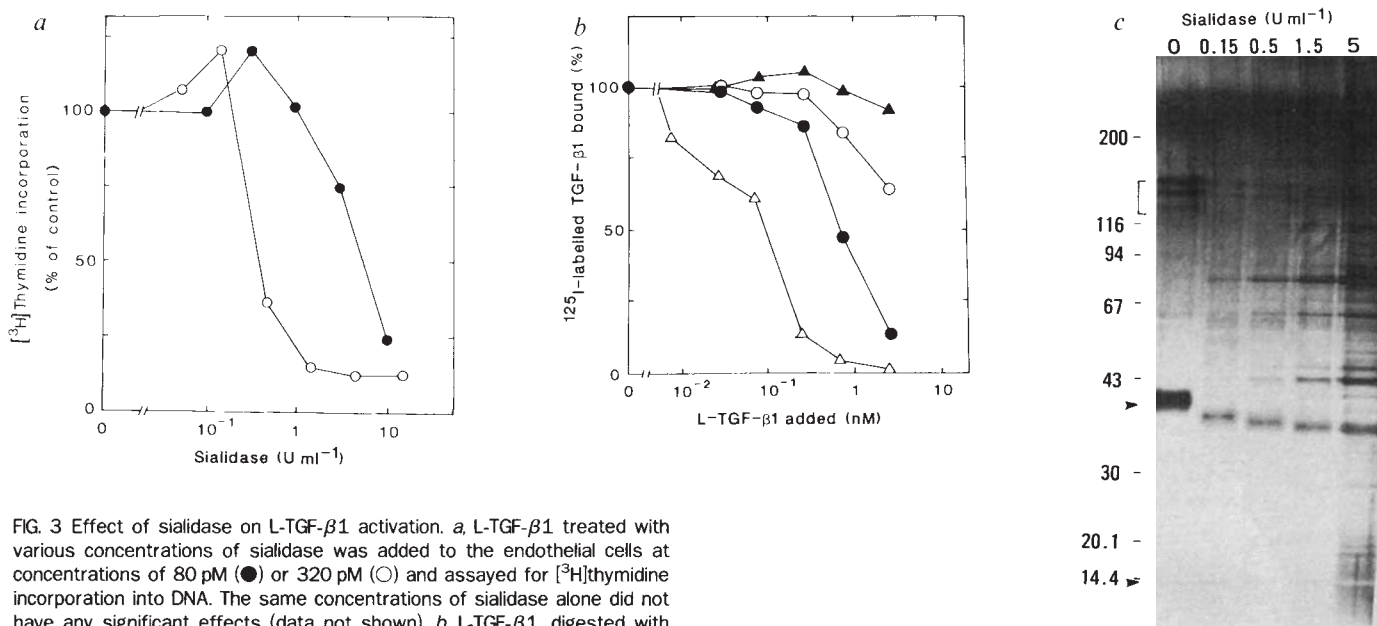


FIG. 3 Effect of sialidase on L-TGF- $\beta$ 1 activation. *a*, L-TGF- $\beta$ 1 treated with various concentrations of sialidase was added to the endothelial cells at concentrations of 80 pM (●) or 320 pM (○) and assayed for [ $^3$ H]thymidine incorporation into DNA. The same concentrations of sialidase alone did not have any significant effects (data not shown). *b*, L-TGF- $\beta$ 1, digested with 15 U ml $^{-1}$  of sialidase, was analysed at different concentrations in the radioreceptor assay on NRK-49F cells (●). As controls were tested L-TGF- $\beta$ 1 alone (▲) and L-TGF- $\beta$ 1 mixed with sialidase just before addition to the cells (○). The effect of transiently acidified L-TGF- $\beta$ 1 is also shown (△). *c*, L-TGF- $\beta$ 1 treated with various concentrations of sialidase was analysed by SDS-gel electrophoresis and silver staining. Bracket on the left, TGF- $\beta$ 1-BP. Arrowheads on the left indicate the positions of TGF- $\beta$ 1 and the N-terminal remnant

acid and mannose-6-phosphate activated L-TGF- $\beta$ 1 at high concentrations: the incorporation of [ $^3$ H]thymidine was 58% of control levels in the presence of 15 mM sialic acid and 42% with 50 mM mannose-6-phosphate. Closely related components such as mannose and mannose-1-phosphate, other monosaccharides such as fucose, *N*-acetylglucosamine and glucuronic acid, and also aspartic acid were all less effective. In the radioreceptor assay also, sialic acid and mannose-6-phosphate significantly activated L-TGF- $\beta$ 1.

To investigate the possibility that sialic acid residues in the precursor remnant are involved in binding TGF- $\beta$ 1 in the latent complex, we used sialidase, which specifically removes sialic acid residues<sup>18</sup>. Sialidase activated L-TGF- $\beta$ 1 in a dose-dependent way, in the growth inhibition assay (Fig. 3*a*). L-TGF- $\beta$ 1 treated with 15 U ml $^{-1}$  sialidase gave half-maximal competition in the binding assays at 0.5 nM, a 40-fold lower concentration than for untreated L-TGF- $\beta$ 1 (Fig. 3*b*), although sialidase could have a direct effect on  $^{125}$ I-labelled TGF- $\beta$ 1 binding (Fig. 3*b*). The 40K band of the precursor remnant on SDS gels was progressively and quantitatively degraded to 36K (Fig. 3*c*). An increased mobility on SDS-gels, also observed using recombinant material<sup>16</sup>, is an expected result of removal of sialic acid residues.

Analyses of recombinant TGF- $\beta$ 1 showed that the carbohydrate structures in the precursor remnant were endo H-resistant, complex oligosaccharides containing sialic acid<sup>16</sup>. In addition, mannose-6-phosphate is present in two of the three glycosylation sites of the precursor remnant<sup>19</sup>, although mannose-6-phosphate is usually found in oligomannose-type oligosaccharides<sup>20,21</sup>, but we do not know whether mannose-6-phosphate is present in the L-TGF- $\beta$ 1 purified from human platelets.

It has been reported that some of the latent TGF- $\beta$  produced by fibroblasts was activated by treatment with proteases, such as plasmin and cathepsin D<sup>12</sup>. L-TGF- $\beta$ 1 from human platelets was, however, not activated by treatment with several proteases (data not shown). If removal of the carbohydrate is part of the mechanism of activation of L-TGF- $\beta$ 1 *in vivo*, how could it be achieved? There are several possible mechanisms. Enzymes

of the TGF- $\beta$ 1 precursor. Arrowhead on the right, the degraded form of the precursor remnant. In these experiments, type V sialidase (Sigma) was used. Very pure sialidase (type X; Sigma) gave essentially the same results (data not shown). Relative molecular masses are shown on the left, in thousands. METHODS. L-TGF- $\beta$ 1 was incubated with sialidase (from *Clostridium perfringens*; Sigma) in 100 mM sodium acetate buffer, pH 5.6, for 18 h at 37 °C.

could remove or modify sialic acid or mannose-6-phosphate residues. Interestingly, it has been shown that sialidase activity can be induced, for example in activated macrophages<sup>22,23</sup>. Alternatively, low relative molecular mass substances produced locally could displace TGF- $\beta$ 1 from the carbohydrate structures of L-TGF- $\beta$ 1. Finally, an interaction between L-TGF- $\beta$ 1 and mannose-6-phosphate receptors could activate TGF- $\beta$ 1. The mannose-6-phosphate receptor has recently been shown to be identical to the insulin-like growth factor II receptor<sup>24</sup>. Future studies will investigate whether any of these mechanisms are involved in regulation of the activity of TGF- $\beta$ 1 and other members of the TGF- $\beta$  family. □

Received 29 December 1988; accepted 2 February 1989.

- Sporn, M. B., Roberts, A. B., Wakefield, L. M. & de Crombrughe, B. *J. Cell Biol.* **105**, 1039–1045 (1987).
- Moses, H. L. *et al.* in *Growth Factors and Transformation*. Cancer Cells 3 (eds Feramisco, J., Ozanne, B. & Stiles, C.) 65–71 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1985).
- Massagué, J. *Cell* **48**, 437–438 (1987).
- Derynck, R. *et al.* *Nature* **316**, 701–705 (1985).
- Cheifetz, S. *et al.* *Cell* **48**, 409–415 (1987).
- de Martin, R. *et al.* *EMBO J.* **6**, 3673–3677 (1987).
- ten Dijke, P., Hansen, P., Iwata, K. K., Pieler, C. & Foulkes, J. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4715–4719 (1988).
- Derynck, R. *et al.* *EMBO J.* **7**, 3737–3743 (1988).
- Pircher, R., Lawrence, D. A. & Jullien, P. *Cancer Res.* **44**, 5538–5543 (1984).
- Pircher, R., Jullien, P. & Lawrence, D. A. *Biochem. biophys. Res. Commun.* **136**, 30–37 (1986).
- Miyazono, K., Hellman, U., Wernstedt, C. & Heldin, C.-H. *J. Biol. Chem.* **263**, 6407–6415 (1988).
- Lyons, R. M., Keski-Oja, J. & Moses, H. L. *J. Cell Biol.* **106**, 1659–1665 (1988).
- Wakefield, L. M., Smith, D. M., Flanders, K. C. & Sporn, M. B. *J. Biol. Chem.* **263**, 7646–7654 (1988).
- Massagué, J. *Meth. Enzym.* **146**, 174–195 (1987).
- Massagué, J. & Like, B. *J. Biol. Chem.* **260**, 2636–2645 (1985).
- Brunner, A. M., Gentry, L. E., Cooper, J. A. & Purchio, A. F. *Molec. cell Biol.* **8**, 2229–2232 (1988).
- Elder, J. H. & Alexander, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4540–4544 (1982).
- Thotakura, N. R. & Bahl, O. P. *Meth. Enzym.* **138**, 350–359 (1987).
- Purchio, A. F. *et al.* *J. Biol. Chem.* **263**, 14211–14215 (1988).
- Hasilik, A. & Neufeld, E. F. *J. Biol. Chem.* **255**, 4946–4950 (1980).
- Tabas, I. & Kornfeld, S. *J. Biol. Chem.* **255**, 6633–6639 (1980).
- Shauer, R. *Adv. Carbohydr. Chem.* **40**, 131–234 (1982).
- Pilatte, Y., Bignon, J. & Lambre, C. R. *Biochim. biophys. Acta* **923**, 150–155 (1987).
- Morgan, D. O. *et al.* *Nature* **329**, 301–307 (1987).
- Frolik, C. A., Wakefield, L. M., Smith, D. M. & Sporn, M. B. *J. Biol. Chem.* **259**, 10995–11000 (1984).
- Blöbel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
- Morrissey, J. H. *Analyt. Biochem.* **117**, 307–310 (1981).

ACKNOWLEDGEMENTS. We thank Linda Baltell for valuable help in the preparation of this manuscript, and Drs Bengt Westermark, Lena Claesson-Welsh, Masaki Saito and Fumimaro Takaku for useful discussions.



# Origin of luteinizing hormone-releasing hormone neurons

Marlene Schwanzel-Fukuda & Donald W. Pfaff

Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

NEURONS expressing luteinizing hormone-releasing hormone (LHRH), found in the septal-preoptic nuclei and hypothalamus<sup>1-3</sup>, control the release of gonadotropic hormones from the anterior pituitary gland<sup>4,5</sup> and facilitate reproductive behaviour<sup>6,7</sup>. LHRH-expressing neurons are also found in the nervus terminalis<sup>8-11</sup>, a cranial nerve that is a part of the accessory olfactory system and which projects directly from the nose to the septal-preoptic nuclei in the brain<sup>12,13</sup>. During development, LHRH-immunoreactivity is detected in the peripheral parts of the nervus terminalis before it is found in the brain<sup>8,10,11</sup>. Using a combination of LHRH immunocytochemistry and tritiated thymidine autoradiography in

fetal mice, we show that LHRH neurons originate in the medial olfactory placode of the developing nose, migrate across the nasal septum and enter the forebrain with the nervus terminalis, arching into the septal-preoptic area and hypothalamus. Clinically, this migratory route for LHRH-expressing neurons could explain the deficiency of gonadotropins seen in 'Kallmann's syndrome' (hypogonadotropic hypogonadism with anosmia).

Experiments were performed on 38 Swiss mice from mothers that had been injected with tritiated thymidine ( $5 \mu\text{Ci g}^{-1}$  of body weight) on days 9, 10, 11, 12, 13, 14, 16, 18, 19 or 20 of gestation. One hour later, these animals were anaesthetized with Metofane, killed by rapid decapitation, laparotomized, and the embryos quickly removed from the uterus and fixed by immersion in Bouin's solution, as preparation for quantitative microscopy. Slides from the areas of the head that included the olfactory placode, nervus terminalis or preoptic area were processed for the immunocytochemical localization of LHRH. Alternate slides were also processed for autoradiographic visualization of the tritiated thymidine.

Luteinizing - hormone - releasing - hormone - immunoreactive

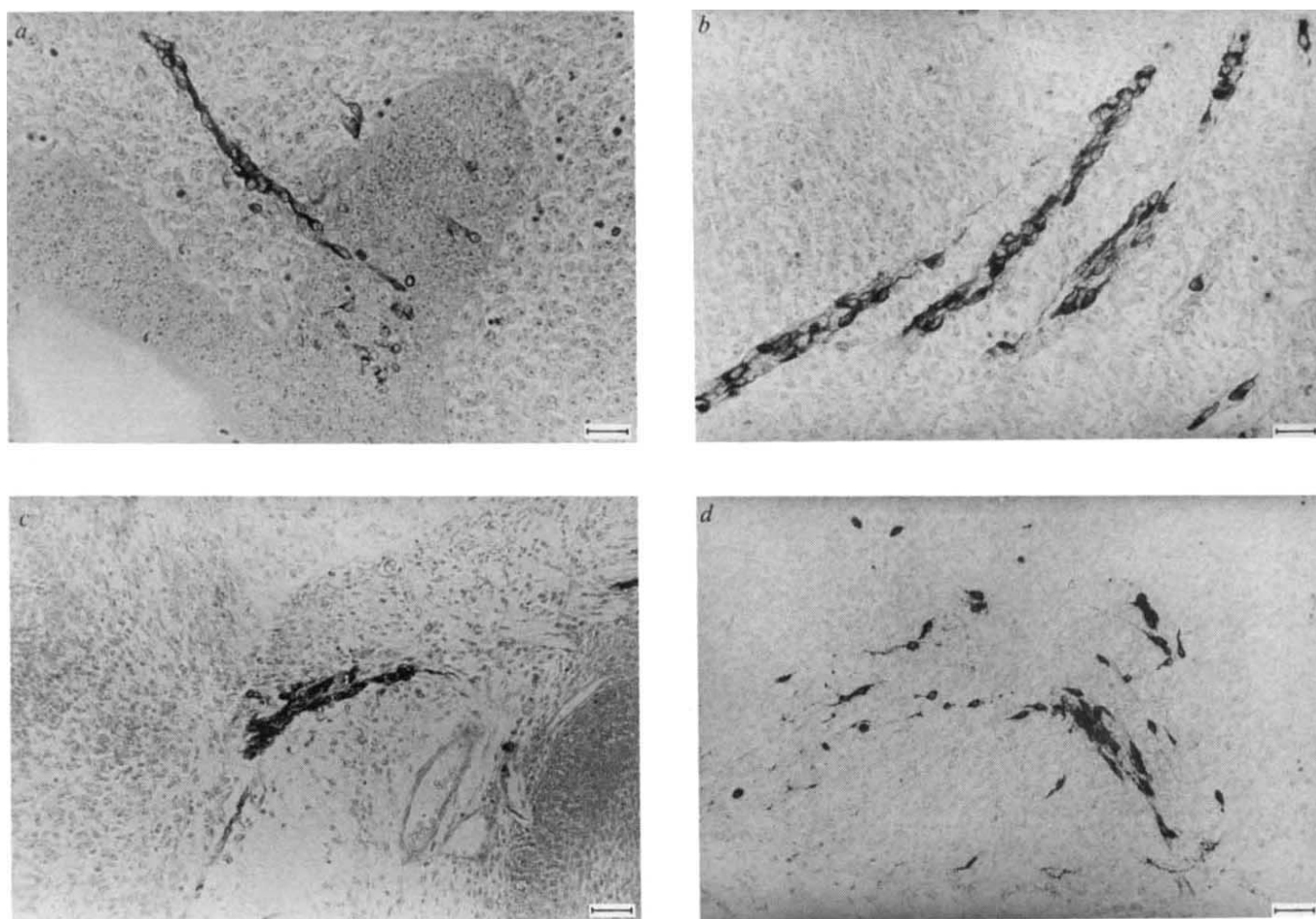


FIG. 1 LHRH-immunoreactive neurons in the embryonic mouse. *a*, At embryonic day 11 LHRH neurons are seen in the epithelium of the primordium of the vomeronasal organ and medial wall of the olfactory pit. Some LHRH cells are seen migrating out of the epithelia with axons of the nervus terminalis and vomeronasal nerves. Scale bar,  $15 \mu\text{m}$ . *b*, Cords of LHRH neurons on the nasal septum at embryonic day 13. Scale bar,  $15 \mu\text{m}$ . *c*, LHRH neurons in the ganglion terminale, caudal to the anlage of the olfactory bulbs, embryonic day 14. Scale bar,  $45 \mu\text{m}$ . *d*, LHRH neurons arch into the septal-preoptic area of the brain at embryonic day 16. Scale bar,  $45 \mu\text{m}$ .

**METHODS.** The antiserum to LHRH, LR1 A-LHRH, kindly provided by Dr Robert Benoit, was raised in a rabbit against [D-Lys6]-luteinizing hormone-releasing hormone conjugated to ovalbumin using glutaraldehyde. The antigenic determinant consists of amino acids 2-4 and 7-10 of the decapeptide and the

antiserum does not crossreact with thyrotropic-releasing factor, substance P, arginine vasopressin, somatostatin-14, met-enkephalin or leu-enkephalin at concentrations of up to  $1 \mu\text{g ml}^{-1}$ . The immunocytochemical localization of LHRH was performed as previously described<sup>10,11</sup>, with unlabelled primary antibody, biotinylated secondary antibody and an avidin-biotin-horseradish peroxidase complex (Vectastain 'ABC' Elite technique, Vector Labs) with 3,3'-diaminobenzidine tetrahydrochloride (DAB, Sigma) as the chromagen. LHRH immunoreactivity was seen as dark brown granular reaction product in the cytoplasm and neurites of certain neurons. To demonstrate the specificity of the antiserum to LHRH, adjacent sections were incubated with antiserum that had been absorbed with  $1 \mu\text{g}$  of synthetic LHRH (Bachem Laboratories) per ml of LR1 A-LHRH diluted 1:10,000 for 24 hours before use. These control sections showed no immunoreactivity.



FIG. 2 The migratory route of LHRH-immunoreactive neurons from derivatives of the medial olfactory placode to the forebrain is shown in microprojection drawings in the sagittal plane of 6- $\mu$ m sections through the whole heads of fetal mice on embryonic (E) days 11, 13, 14 and 16. The black dots represent LHRH-immunoreactive neurons. Abbreviations: gt, ganglion terminale; ob, olfactory bulb; poa, preoptic area; vno, vomeronasal organ. On day 11, LHRH-immunoreactive cells are seen in the anlage of the vomeronasal organ and medial wall of the olfactory placode. On day 13, most of the LHRH cells are seen on the nasal septum, with the nervus terminalis and the vomeronasal nerves and on day 14, most of these cells are in the ganglion terminale and in the central roots of the nervus terminalis. The 16-day-old fetal brain had most LHRH neurons arching through the forebrain into the hypothalamus and preoptic areas.

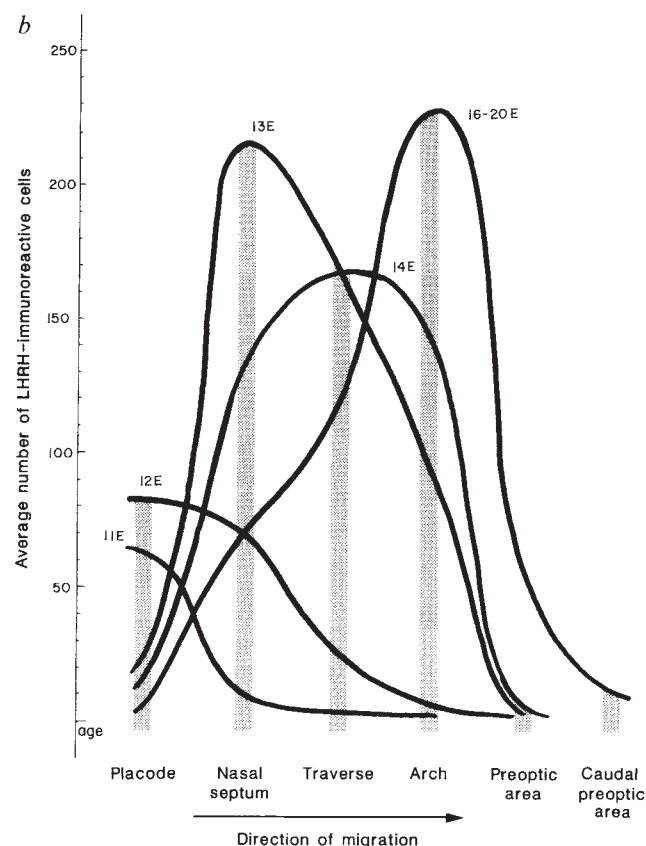
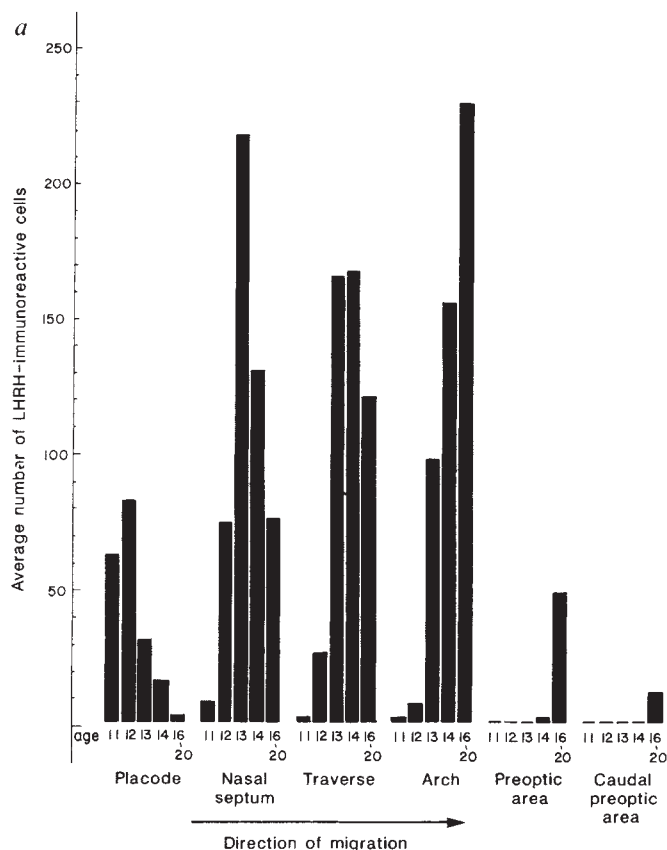
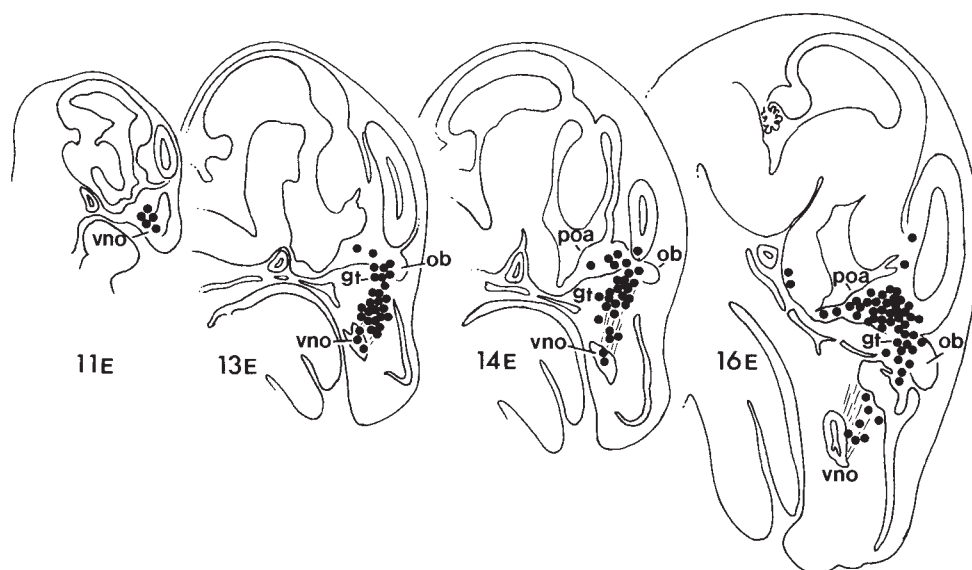


FIG. 3 *a*, The distribution of average numbers of LHRH-immunoreactive cells counted on embryonic days 11 ( $n=6$ ), 12 ( $n=3$ ), 13 ( $n=4$ ), 14 ( $n=5$ ), 16 ( $n=3$ ), 18 ( $n=3$ ), 20 ( $n=2$ ) are plotted to test the proposed migratory route. We have classified the neuroanatomical areas of the brain and nose under general headings. Placode, epithelia of the olfactory placode, the olfactory pit and the anlage of the vomeronasal organ. Nasal septum, the area on the nasal septum below the cribriform plate of the ethmoid bone. Traverse, the area above the cribriform plate, ventromedial to the olfactory bulbs or their anlagen, including the ganglion terminale and intracranial roots of the nervus terminalis. Arch, the central roots of the nervus terminalis, the telencephalic vesicle of the 11 to 14-day-old embryos and the olfactory tubercle, the anterior olfactory nucleus, the nucleus and tract of the diagonal band, the medial septal nucleus and the anterior hypothalamus of older

fetuses. Preoptic area, both the medial and lateral preoptic nuclei. Caudal preoptic area, the organum vasculosum of the lamina terminalis and the areas dorsal and caudal to the optic chiasm. Each age group has a characteristic pattern of distribution of LHRH-immunoreactive neurons. Any sample of the distribution of LHRH cells in the embryonic mouse fitted into one of these curves with less than a one-day difference ( $\pm 1$  day) from embryonic day 11 to day 14. After 16 days of gestation most migration seems over. As little difference in the distribution of LHRH-immunoreactive neurons was found between days 16, 18, 19 and 20 of embryonic life these ages were grouped together. *b*, Schematic summary of *a*. Vertical bars illustrate where we found the most cells at a given age and the dark lines joining the bars illustrate the 'envelope' characteristic of a given age.

cells were discovered in the medial olfactory placode of fetal mice. Their subsequent positions, described as a function of developmental age, form a coherent migratory route across the nasal septum, entering the forebrain with the nervus terminalis, coursing medial to the olfactory bulbs and arching into the septal-preoptic region of the brain.

By the 11th day of embryonic age, LHRH immunoreactivity was detected in the cytoplasm of several cells in the epithelium of the medial part of the olfactory placode, including the part that forms the anlage of the vomeronasal organ (Fig. 1a and Fig. 2). These cells were oval or fusiform in shape, oriented perpendicular to the apical (luminal) surface and seen in the apical, intermediate and basal layers of the epithelium. Occasionally, a single process of an LHRH-immunoreactive cell body in the basal surface extended to the apical surface. No LHRH immunoreactivity was found in the medial olfactory pit before formation of the primordium of the vomeronasal organ. In some animals LHRH-immunoreactive cells migrated out of the epithelium of the medial wall of the olfactory pit, with axons of the nervus terminalis and/or vomeronasal nerves (Fig. 1a). Through serial sagittal sections, LHRH-immunoreactive cells were traced into the primordium of the principal ganglion of the nervus terminalis. The nervus terminalis and vomeronasal nerves were continuous with the forebrain, medial to the ingrowing axons of the olfactory nerves. No LHRH-immunoreactive neurons were observed in the forebrain if these structures, which form a bridge between the nasal cavity and the forebrain, were not present.

At 12 and 13 days of gestation, relatively large numbers of LHRH-immunoreactive neurons were seen in the epithelia of the placode-derived medial olfactory pit and vomeronasal organ, and cords of LHRH-immunoreactive cells coursed from these structures across the nasal septum (Fig. 1b and Fig. 2) and traversed the ganglion terminale and central roots of the nervus terminalis into the ventromedial forebrain (Fig. 1c and Fig. 2). From this point, most LHRH immunoreactive cells formed an arch through the developing forebrain (Fig. 1d) into the anlagen of the septum and hypothalamus, extending into the preoptic area by embryonic day 14 (Fig. 2).

The numbers of LHRH-immunoreactive neurons along this proposed migratory route support the notion of migration from a medial olfactory pit origin. With increasing age, from days 16 to 20 of gestation, and in the newborn mouse, an increase in the number of LHRH-immunoreactive cells was found in the septum, the hypothalamus and the preoptic area, while a concomitant decrease was observed in the number of LHRH-immunoreactive cells seen in the epithelia of the vomeronasal organ and medial wall of the olfactory pit and on the nasal septum (Fig. 3).

Experiments using a combination of tritiated thymidine autoradiography with LHRH-immunocytochemistry failed to provide support for an alternative hypothesis: that the neurons originate from the surface of the ventricles. The presence of the preoptic LHRH neurons can be accounted for by the proposed migration route from the olfactory placode. Although tritiated thymidine autoradiography showed cells proliferating and moving out from the ependymal surface of the lateral ventricles of the forebrain no labelled cells in this area expressed LHRH. If preoptic LHRH neurons had followed the short route from the lateral or third ventricle, labelled cells would have been found here. Instead, a few LHRH-immunoreactive neurons on the nasal septum, just outside the epithelium of the olfactory pit, had labelled nuclei (Fig. 4).

We concluded that LHRH neurons originate in the epithelia of the vomeronasal organ and medial wall of the olfactory pit, and enter the forebrain with branches of the nervus terminalis and vomeronasal nerves. An olfactory epithelial origin for LHRH neurons provides an explanation for the puzzling observation that some of them have cilia<sup>14,15</sup>.

Clinical studies have shown that gonadal dysplasia with

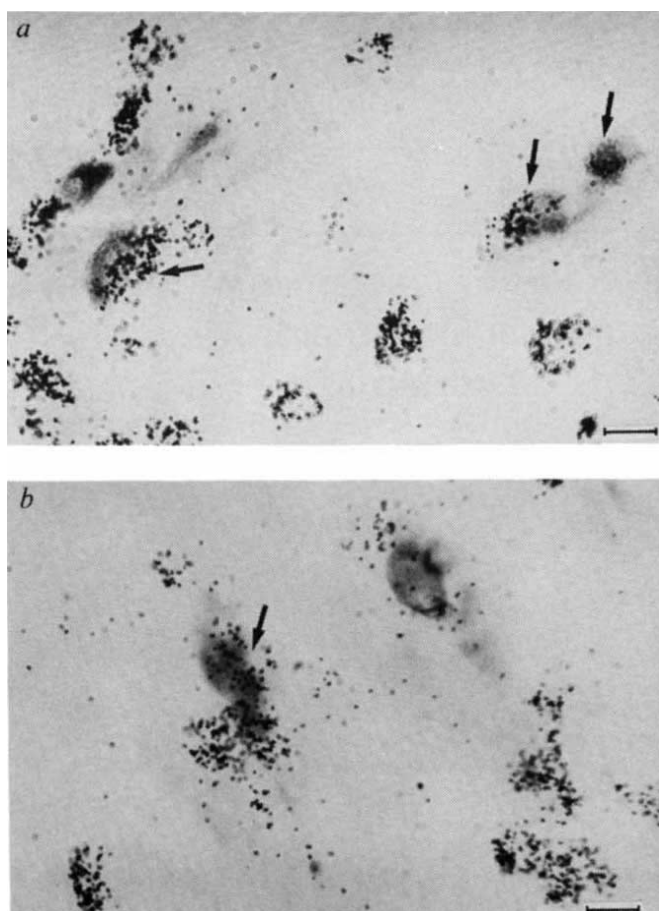


FIG. 4 *a*, Three LHRH immunoreactive neurons, with overlying black autoradiographic silver grains indicating the incorporation of tritiated thymidine into their nuclei (arrows), are seen on the nasal septum just outside of the olfactory pit (lower left) of a 12-day-old fetal mouse. An unlabelled LHRH neuron, with a clear nucleus, is seen in the upper left-hand corner adjacent to labelled cells. *b*, One LHRH-immunoreactive neuron with a thymidine-labelled nucleus (arrow) is seen adjacent to a non-immunoreactive autoradiographically-labelled cell nucleus, on the nasal septum outside of the anlage of the vomeronasal organ (lower right), of a 13-day-old fetal mouse. An unlabelled LHRH neuron is seen on the upper right. Scale bar, 10  $\mu$ m.

decreased secondary sex characteristics and deficiency in gonadotropins can be associated with malformation of the olfactory apparatus. A study by Kallmann<sup>16</sup> of the familial occurrence of anosmia associated with gonadal failure in human males ascribes a genetic basis to hypogonadotropic hypogonadism with anosmia, which is now known as Kallmann's syndrome. The hereditary syndrome is more common in males but may be present in females as well. As the anosmia of Kallmann's syndrome is associated with agenesis of the olfactory bulbs, whose development depends upon contact with the olfactory nerves, the failure of axonal and LHRH neuronal migration from the olfactory placode would explain both the anosmia and the gonadotropic deficiencies of this syndrome. □

Received 12 October 1988; accepted 3 January 1989.

1. Krey, L. C. & Silverman, A.-J. in *Brain Peptides* (eds Kreiger, D. T., Brownstein, M. J. & Martin, J. B.) 687–709 (Wiley New York, 1983).
2. Shivers, B. D., Harian, R. E., Morrell, J. E. & Pfaff, D. W. *Neuroendocrinol.* **36**, 1–12 (1983).
3. Silverman, A.-J. & Zimmerman, E. A. in *Brain-Endocrine Interaction III. Neural Hormones and Reproduction*, 83–96 (Karger, Basle, 1978).
4. Schally, A. V. et al. *Biochem. biophys. Res. Commun.* **43**, 393–399 (1971).
5. Guillemin, R. *Science* **202**, 390–402 (1978).
6. Pfaff, D. W. *Science* **182**, 1148–1149 (1973).
7. Moss, R. L. & McCann, S. M. *Science* **181**, 177–179 (1973).
8. Schwanzel-Fukuda, M. & Silverman, A.-J. *J. comp. Neurol.* **191**, 213–225 (1980).
9. Jennes, L. & Stumpf, W. E. *Cell Tissue Res.* **209**, 239–256 (1980).
10. Schwanzel-Fukuda, M., Morrell, J. I. & Pfaff, D. W. *J. comp. Neurol.* **238**, 348–364 (1985).
11. Schwanzel-Fukuda, M., Fadem, B. H., Garcia, M. S. & Pfaff, D. W. *J. comp. Neurol.* **276**, 44–60 (1988).



12. Larsell, O. *Ann. Otol. Rhinol. Lar.* **59**, 414–435 (1950).
13. Pearson, A. A. *J. comp. Neurol.* **75**, 39–66 (1941).
14. Kozlowski, G. P., Chu, L., Hostetter, G. & Kerdellhue, B. *Peptides* **1**, 37–46 (1980).
15. Jennes, L., Stumpf, W. E. & Sheely, M. J. *J. comp. Neurol.* **232**, 534–547 (1985).
16. Kallman, F. J., Schoenfeld, W. A. & Barrera, S. E. *Am. J. ment. Deficiency* **48**, 203–236 (1944).

ACKNOWLEDGEMENTS. We thank Haruto Fukuda for preparation of the graph, Steven Kleopoulos for technical assistance and Adelaide Acquaviva for preparation of the manuscript. This research was supported by a grant from the NIH and funds from the Whitehall Foundation.

## Effects of modulators of myosin light-chain kinase activity in single smooth muscle cells

Takeo Itoh, Mitsuo Ikebe\*, Gary J. Kargacin, David J. Hartshorne†, Bruce E. Kemp‡ & Fredric S. Fay

Department of Physiology, University of Massachusetts Medical Center, 55 Lake Avenue North, Worcester, Massachusetts 01655, USA

\* Department of Physiology, Case-Western Reserve University, Cleveland, Ohio 44106, USA

† Department of Nutrition and Food Science, University of Arizona, 309 Agricultural Science Building, Tucson, Arizona 85721, USA

‡ St Vincent's Institute of Medical Research, Melbourne 3065, Australia

**PHOSPHORYLATION of myosin light chains by a calmodulin-myosin light-chain kinase (MLCK) pathway is considered to be responsible for coupling increased calcium concentration with contraction in smooth muscle<sup>1,2</sup>. This simple view has, however, recently been questioned<sup>3–6</sup>. To test this hypothesis directly, we microinjected individual smooth muscle cells with modulators of the MLCK pathway while measuring contraction and calcium-ion concentration. Injection of a constitutively active proteolysed form of MLCK causes contraction but no change in calcium concentration. By contrast, injection of peptide inhibitors of MLCK blocks contraction in response to K<sup>+</sup> depolarization, despite the fact that the change in calcium concentration in response to stimulation was enhanced over controls. These results provide a direct demonstration at the level of a single cell that activation of the calmodulin-MLCK pathway is both necessary and sufficient to trigger contraction of smooth muscle.**

In these experiments we used single smooth-muscle cells from the stomach muscularis of the toad *Bufo marinus*<sup>7</sup>. These cells have been well characterized<sup>8–10</sup>, which has helped us to assess the effect of modulators on the contraction process. We microinjected the modulators under pressure into single cells through a glass micropipette, and the contractile activity was then recorded by video using phase-contrast images of the cells at 680 nm illumination. Effects on [Ca<sup>2+</sup>] were simultaneously determined (see Fig. 1a) by monitoring 510 nm fluorescence of fura 2 at two excitation wavelengths (340 and 380 nm) using a high-speed digital dual-wavelength microfluorimeter<sup>8,10,11</sup>.

We first confirmed that the toad stomach cells contained a system that would respond to myosin light-chain phosphorylation, and also established that purified turkey gizzard MLCK could effectively phosphorylate toad myosin light chains. We found that actomyosin prepared from toad stomach had a basal ATPase activity of 0.5 nmol min<sup>-1</sup> per mg protein, but could be stimulated to 24 nmol min<sup>-1</sup> per mg protein by adding Ca<sup>2+</sup>-calmodulin and either gizzard MLCK or toad-stomach MLCK. The increase in myosin light-chain phosphorylation and actomyosin ATPase activity could be induced independently of Ca<sup>2+</sup> by a tryptic fragment (molecular weight 61,000; 61K) of MLCK (IMLCK) from turkey gizzard; this fragment is constitutively active as a result of the removal of an inhibitory portion of the molecule<sup>12,13</sup>. The activity of toad-stomach actomyosin ATPase

is increased by IMLCK from 0.5 to 25 nmol min<sup>-1</sup> per mg protein and myosin light chain phosphorylation is increased from 0 to 1.9 mol phosphate per mol myosin in the presence of 1 mM EGTA. An autoradiograph of toad-stomach actomyosin phosphorylated both by the native MLCK and by IMLCK showed that only the 20K myosin light chains were phosphorylated. These results indicate that the actomyosin ATPase from the toad stomach is subject to control by Ca<sup>2+</sup> through an MLCK pathway analogous to that found in other smooth muscles.

We developed procedures for the microinjection of single smooth muscle cells with modulators of this pathway to investigate their role in the regulation of contraction. As illustrated in Fig. 1b, insertion of the microinjection pipette into cells caused a small transient contraction, presumably as a result of a momentary leakage of ions around the micropipette. Subsequent injection of buffer into the cells (2.2 ± 0.7% of the cell volume; *n* = 14) did not cause a significant contraction of the cells (length was 97.8 ± 3.5% of the length before injection; *n* = 4) or a significant change in [Ca<sup>2+</sup>] 138.2 ± 17.2 nM after injection (*n* = 4) versus 143.4 ± 11.2 nM in impaled, non-injected cells (*n* = 22). As can be seen in Fig. 1c, compounds injected into the cells equilibrate rapidly. In this case, injected fura 2 was equilibrated to 50% by four seconds after the start of the injection.

The effect of injection of the IMLCK on contractile activity and [Ca<sup>2+</sup>] is shown in Fig. 2. It can be seen that injection of this independent kinase caused marked cell-shortening without a significant change in [Ca<sup>2+</sup>]. No effect on either parameter was seen in four cells following injection of comparable amounts of boiled IMLCK. The magnitude and maximum rate of shortening observed in response to injection of IMLCK was virtually identical to that seen following micro-injection of Ca<sup>2+</sup> or depolarization by K<sup>+</sup>-substituted amphibian physiological salt solution. Brief high K<sup>+</sup>-depolarization and Ca<sup>2+</sup> injection produced a transient rise in [Ca<sup>2+</sup>] and a cell-shortening that reversed after [Ca<sup>2+</sup>] had returned to rest (Fig. 2a, inset). By contrast, the contraction induced by injection of IMLCK did not reverse, presumably because intracellular kinase activity remained high. The estimated levels of kinase activity that resulted from typical injections of the purified IMLCK (0.16 nmol phosphate min<sup>-1</sup> per ml cell water) are comparable with the endogenous MLCK activity expected when these cells are maximally activated (0.32 nmol phosphate min<sup>-1</sup> per ml cell water). Hence the effects we observed are occurring under almost physiological conditions. These results indicate that activation of MLCK in the intact smooth muscle cell is itself sufficient to produce a contraction indistinguishable from that seen following stimuli acting to increase [Ca<sup>2+</sup>] in the normal way.

To understand further the mechanism underlying the activation of contraction by both Ca<sup>2+</sup> and MLCK, we tested the effects of two synthetic peptides. Figure 3 shows the effect of co-injection of these peptides with IMLCK. It can be seen that the pseudosubstrate peptide SM1, which corresponds to amino-acid residues 480–501 of the smooth muscle MLCK and has strong substrate-antagonist and weaker calmodulin-antagonist properties<sup>12–14</sup>, blocks the contractile effect of injection of the IMLCK. By contrast, the calmodulin-binding peptide RS20, which corresponds to residues 493–512 and has strong calmodulin-antagonist and extremely weak substrate-antagonist activity<sup>15,16</sup>, has no significant effect on the contraction induced by the IMLCK when co-injected at the same concentration as the SM1 peptide.

We also checked the effects of these two peptides on contraction mediated by changes in [Ca<sup>2+</sup>]. To ensure that the peptide was equilibrated throughout the cell, cytosolic [Ca<sup>2+</sup>] was only increased by external application of a high [K<sup>+</sup>] 5 min after peptide injection. Transient depolarization of a cell with high [K<sup>+</sup>] in both the control and SM1-injected cells caused a transient increase in [Ca<sup>2+</sup>] (Fig. 4a). Although the [Ca<sup>2+</sup>] increased more in SM1-injected cells, the contractile response was typically much smaller than that in control cells. The cal-



modulin-antagonist peptide RS20 also inhibited the contractile response to  $K^+$ -depolarization, despite the fact that the  $Ca^{2+}$  transient was again larger than in control cells. The inhibitory effect of these peptides was evident when we measured either the maximum rate or extent of shortening (Fig. 4c). The inhibitory effect of the RS20 and SM1 peptides on contraction was seen at all calcium concentrations (Fig. 4b). At the concentrations of peptide available for microinjection, neither SM1 nor RS20 fully inhibited the contractile response to  $K^+$  depolarization.

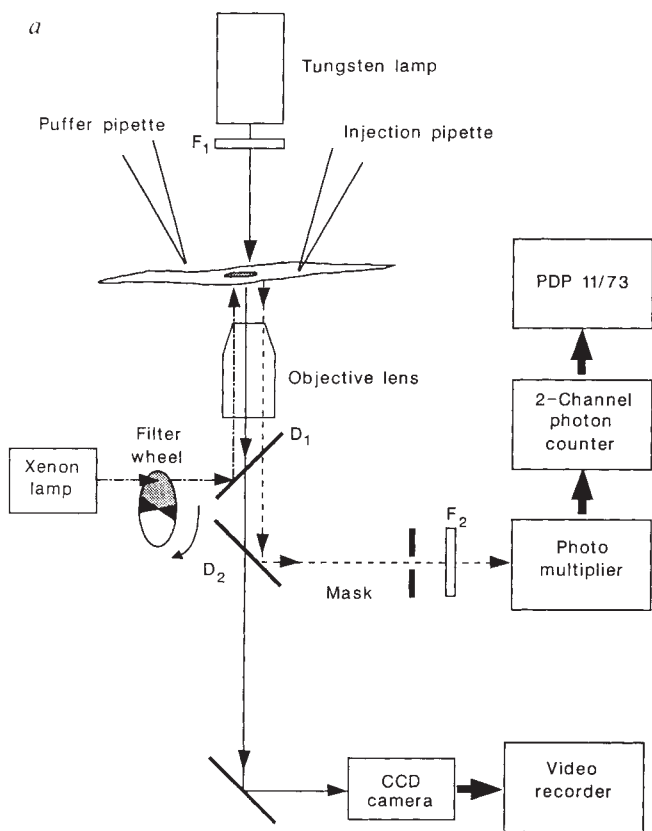
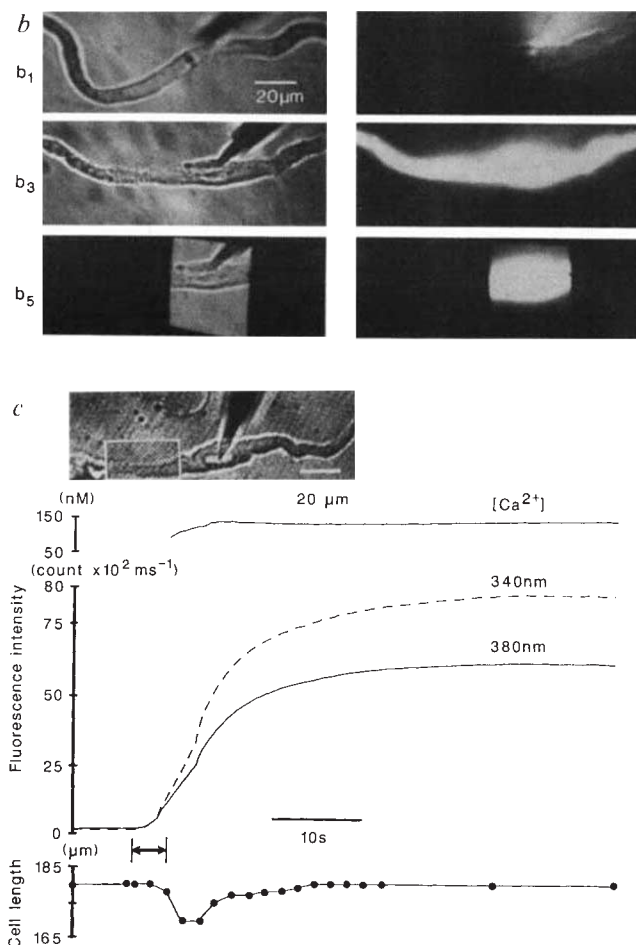


FIG. 1 Assessment of the effect on cell  $[Ca^{2+}]$  and length of injection of solutions into single smooth muscle cells. *a*, Schematic diagram of experimental system. A single cell in a chamber (not shown) filled with amphibian physiological salt solution (APS)<sup>7,17</sup> was impaled with a borosilicate glass micropipette (28–34 M $\Omega$ ; 3 M KCl) and peptides or enzymes dissolved in microinjection buffer (MIB) (0.2 mM EGTA, 20 mM PIPES, pH 6.8) were microinjected under pressure. Cells were stimulated by pressure-ejection of  $K^+$  substituted APS from a puffer pipette (30 K $\Omega$ ; 3 M KCl) within 50  $\mu$ m of the cell. Changes in  $[Ca^{2+}]$  were assessed from measurements of fura-2 fluorescence<sup>18</sup> (dissociation constant,  $K_D$  200 nM<sup>8</sup>) as described previously<sup>8</sup>. Fura-2 fluorescence emission at 510 nm and bright-field images of the cell at 680 nm are separately passed to the photomultiplier and a charged-coupled device camera, respectively, by a dichroic mirror ( $D_2$  reflects light below 640 nm) which was substituted for the photochanger in the IM-35 Zeiss microscope.  $F_2$  and  $F_1$  are interference filters centred at 510 nm (full width at half maximum transmission, 60 nm) and 680 nm (FWHM, 20 nm), respectively. *b*, Images of single smooth-muscle cells during impalement and microinjection. Left, bright-field images of a cell before ( $b_1$ ) and 3 min after impalement ( $b_3$ ) ( $L_{final}$  is 93%  $L_{initial}$ ). Cell after injection with the mask placed in an intermediate image plane to reduce fluorescence background is shown in  $b_5$ . Right, rhodamine fluorescence images of the same cell;  $b_2$ , cell impaled with pipette filled with 10 mM rhodamine-dextran (molecular weight 70,000) from Molecular Probes (Oregon). Image of the cell one min after injection is shown in  $b_6$ ;  $b_6$  shows the image masked as in  $b_5$ . *c*, Calculated  $[Ca^{2+}]$ , fluorescence intensity at 340 and 380 nm, and cell length during injection of fura 2 (5 mM). The recordings were obtained from a masked portion of the cell 20  $\mu$ m away

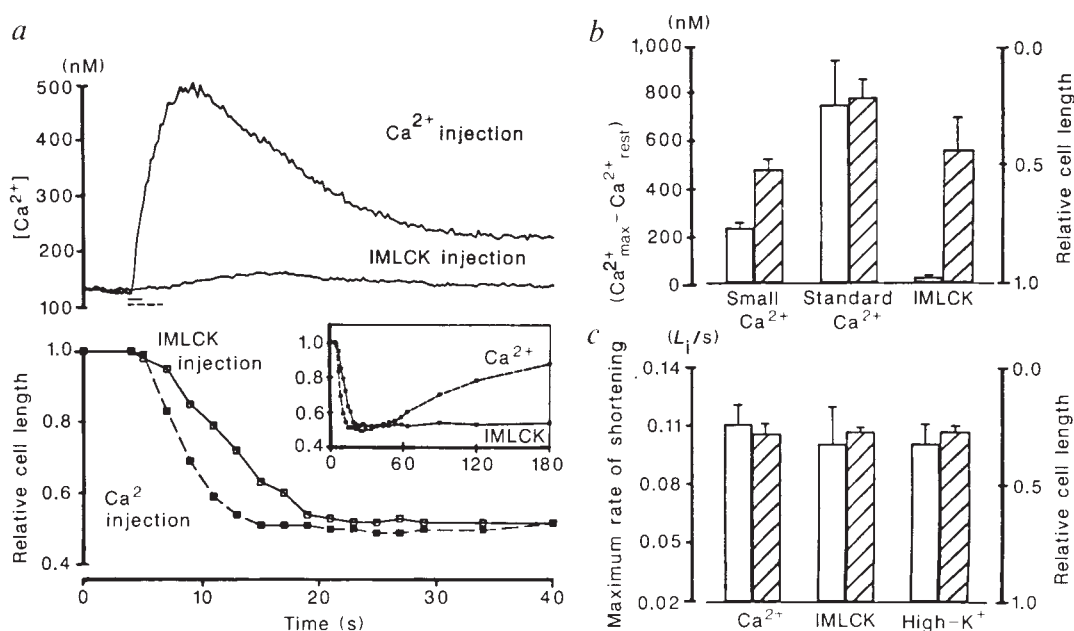
from the point of injection. In skinned cells where higher peptide concentrations were attainable, 0.1 mM SM1 produced virtually complete ( $84 \pm 9\%$ ;  $n = 4$ ) inhibition of the maximum shortening response initiated by 0.1 mM  $Ca^{2+}$  in the presence of 1  $\mu$ M calmodulin and 1 mM ATP. At 10  $\mu$ M SM1, the response of the skinned cells was inhibited by 60%, which is about the level of inhibition of the high  $[K^+]$  response of intact cells injected with SM1 to a concentration of 35  $\mu$ M. The inhibitory effects of the RS20 and SM1 peptides on contraction initiated by  $K^+$  depolarization



from the point of injection.

**METHODS.** Single smooth-muscle cells were isolated by enzyme treatment from *Bufo marinus*<sup>7,17</sup> and loaded with the acetoxy methyl ester of fura 2 (fura 2-AM) before impalement (410 nmol fura 2-AM per  $10^6$  cells; mean intracellular [fura 2],  $34.5 \pm 3.5 \mu$ M). In some experiments fura 2 (free acid) was co-injected with peptides to assess the dilution of material injected into the cell. The dilution factor was calculated by comparing the [fura 2] in the pipette with that in the cell, which was calculated according to the expression<sup>11</sup>  $D = k \cdot [340 FI + C \times 380 FI] / V$ , where 340 FI and 380 FI are the steady-state fluorescent intensities and  $C$  is the ratio at 340 nm/380 nm of the change in the fluorescence of fura 2 going from the  $Ca^{2+}$ -free to the  $Ca^{2+}$ -saturated state;  $V$  is the cell volume contained within the mask, calculated as described in<sup>18</sup>;  $k$  is a proportionality constant relating fluorescent intensity to moles of fura 2, determined by measuring the fluorescence of spherical microdroplets of solutions of known fura 2 concentration in immersion oil. The effects of peptides and enzymes on saponin-skinned cells were determined as described<sup>19</sup>. Myosin and MLCK were purified from toad-stomach muscle and turkey gizzard<sup>12,20</sup>; IMLCK was obtained by tryptic proteolysis of gizzard MLCK and purified by DEAE-ion exchange chromatography<sup>12</sup>. Toad actomyosin or toad myosin plus skeletal-muscle actin (all at 0.2 mg ml<sup>-1</sup>) were incubated at 25 °C in 85 mM KCl, 8 mM MgCl<sub>2</sub>, 30 mM Tris-HCl, (pH 7.5), with different combinations of 10  $\mu$ g ml<sup>-1</sup> MLCK, 10  $\mu$ g ml<sup>-1</sup> IMLCK, 5  $\mu$ g ml<sup>-1</sup> calmodulin, 0.1 mM CaCl<sub>2</sub> and 1 mM EGTA and actomyosin ATPase<sup>21</sup>, and the extent of myosin phosphorylation<sup>22</sup> was measured. The uncertainty in mean values given is the standard error of the mean.

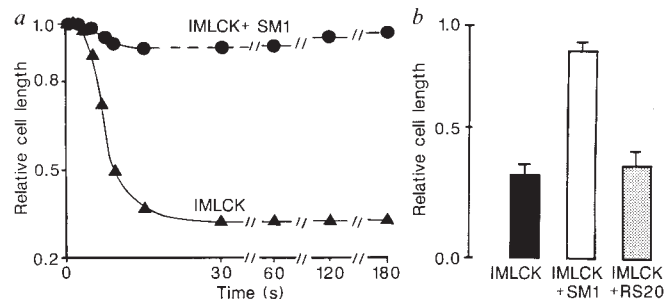
FIG. 2 Effect of microinjection of  $\text{CaSO}_4$  or IMLCK on  $[\text{Ca}^{2+}]$  and the contractile state of single smooth-muscle cells. *a*, Effects on  $[\text{Ca}^{2+}]$  and cell length of injecting  $\text{CaSO}_4$  or IMLCK in two different smooth-muscle cells.  $\text{CaSO}_4$  (0.1 mM) was injected in MIB without EGTA, and IMLCK was injected in MIB at  $1 \text{ mg ml}^{-1}$  (specific activity was  $8 \mu\text{mol phosphate min}^{-1}$  per mg protein at  $25^\circ\text{C}$ , estimated using isolated gizzard myosin light chains as a substrate under the conditions given in the legend to Fig. 1). The short lines below the  $[\text{Ca}^{2+}]$  traces indicate the period of injection of  $\text{CaSO}_4$  (solid) or IMLCK (dashed);  $\text{CaSO}_4$  was injected at 20 psi and the kinase at 40 psi. Cell length is expressed relative to that



before injection. Inset illustrates the changes in cell length over a longer period for the same two cells. *b*, Summary of effects observed on several cells in which  $\text{CaSO}_4$  or IMLCK were injected. 0.1 mM  $\text{CaSO}_4$  was injected according to two different protocols (at 20 psi for 1 s, small  $\text{Ca}^{2+}$  injection; at 40 psi for 3 s, standard  $\text{Ca}^{2+}$  injection). The effects on  $[\text{Ca}^{2+}]$  (open bars) are the peak change in  $[\text{Ca}^{2+}]$  above that before injection. The effect on cell length (cross-hatched bars) is expressed as the peak decrease in cell length after injection, divided by the pre-injection cell length. IMLCK was injected at 40 psi for 3 s. Values shown are the means for 4, 10 and 6 cells (from left to right) under the three conditions. *c*, Effects on the maximum extent (cross-hatched) and maximum rate (open bars) of cell shortening following

$\text{CaSO}_4$  injection, IMLCK injection and extracellular application of  $\text{K}^+$ -substituted APS. The maximum rate of shortening for each cell has been normalized for its length before stimulation. Both the maximum extent and rate of shortening appear to reflect the level of activation, rather than the limits set by the kinetics of the contractile mechanism<sup>19</sup> or the rate at which the excitatory stimulus is delivered to the activation system. This latter point follows from the fact that similar maximum rates of shortening are observed after electrical stimulation using paddles along the cell<sup>23</sup> and when  $\text{Ca}^{2+}$  or IMLCK are injected midway along the cell, despite the fact that the geometry for spread of  $\text{Ca}^{2+}$  or IMLCK is so different. The values plotted are (from left to right) the means for 21, 24 and 14 cells.

FIG. 3 Effect of calmodulin (RS20) and light-chain kinase (SM1) inhibitory peptides on the contractile action of unregulated MLCK (IMLCK). *a*, Effect of peptide SM1 on the contraction induced by IMLCK. IMLCK ( $0.5 \text{ mg ml}^{-1}$ ) with or without SM1 peptide ( $0.5 \text{ mM}$ ) was injected at 60 psi for 3 s into two different cells starting at zero time. *b*, Comparison of the effect of RS20 and SM1 peptides on the contraction induced by the IMLCK. One of the peptides at  $0.5 \text{ mM}$  was co-injected with  $0.5 \text{ mg ml}^{-1}$  IMLCK. Filled, unfilled and hatched bars show the effect of injection of IMLCK without or with the peptides. Data for IMLCK, IMLCK + SM1 and IMLCK + RS20 are for 24, 14 and 8 cells, respectively.



indicate that the pathway by which calcium leads to activation requires both calmodulin and MLCK activation. As  $\text{Ca}^{2+}$ -induced contraction and contraction induced by IMLCK injection can both be fully inhibited by SM1, activation of MLCK must be an essential step in the pathway leading from  $[\text{Ca}^{2+}]$  increases to contraction. It would appear that changes in all contractile properties assessed in these studies in response to  $\text{Ca}^{2+}$  can be attributed to a calmodulin/myosin light chain phosphorylation pathway mechanism.

The peptides RS20 and SM1 also cause significant changes in the  $\text{Ca}^{2+}$  transient induced by  $\text{K}^+$ -depolarization. Figure 4*a* shows that the resting  $[\text{Ca}^{2+}]$  was unchanged, but the response to  $\text{K}^+$  depolarization was enhanced, by both peptides; the peak  $[\text{Ca}^{2+}]$  change was  $329 \pm 28 \text{ nM}$  ( $n = 50$ ),  $668 \pm 154 \text{ nM}$  ( $n = 9$ ) and  $1,223 \pm 294 \text{ nM}$  ( $n = 3$ ) for the control, RS20- and SM1-injected cells, respectively. These peptides also altered the kinetics of the  $[\text{Ca}^{2+}]$  transient. The rate of rise of  $[\text{Ca}^{2+}]$  was enhanced by SM1 but not by RS20: the maximal rate of  $[\text{Ca}^{2+}]$  rise was  $262 \pm 64 \text{ nM s}^{-1}$  ( $n = 3$ ) in SM1-injected cells, and  $82 \pm 37 \text{ nM s}^{-1}$  ( $n = 9$ ) and  $84 \pm 19 \text{ nM s}^{-1}$  ( $n = 16$ ) in RS20 and control cells, respectively. Both RS20 and SM1 slowed the rate of return of  $[\text{Ca}^{2+}]$  to resting levels after brief  $\text{K}^+$  depolarization. For

example, the rate of decline of  $[\text{Ca}^{2+}]$  at  $400 \text{ nM}$   $[\text{Ca}^{2+}]$  was  $-29 \pm 3 \text{ nM s}^{-1}$  ( $n = 4$ ) in control cells and  $-15 \pm 7 \text{ nM s}^{-1}$  ( $n = 3$ ) and  $-19 \pm 1 \text{ nM s}^{-1}$  ( $n = 3$ ) for cells injected with RS20 and SM1, respectively. Finally, the  $[\text{Ca}^{2+}]$  increase in response to brief  $\text{K}^+$  depolarization in RS20-injected cells continued for several seconds after application of  $\text{K}^+$ -substituted amphibian physiological salt solution, an effect which was not seen in control or SM1-injected cells. These observations do not seem to be the result of non-specific effects of these peptides. Injection of control peptides that are the same as SM1 and RS20, apart from a single amino-acid substitution that results in loss of activity towards the calmodulin/MLCK pathway, had no effect on the  $\text{Ca}^{2+}$  transient following  $\text{K}^+$  depolarization. Our results indicate that the calmodulin/MLCK pathway or some unknown calmodulin/protein kinase system plays a direct or indirect role in limiting the size and time course of the  $[\text{Ca}^{2+}]$  change to depolarization. While the observed effects of the modulators of MLCK activity on  $[\text{Ca}^{2+}]$  raise new questions, the observed effects on contraction settle others. The latter results clearly and directly demonstrate that activation on the calmodulin/MLCK pathway is both necessary and sufficient for triggering contraction of smooth muscle.

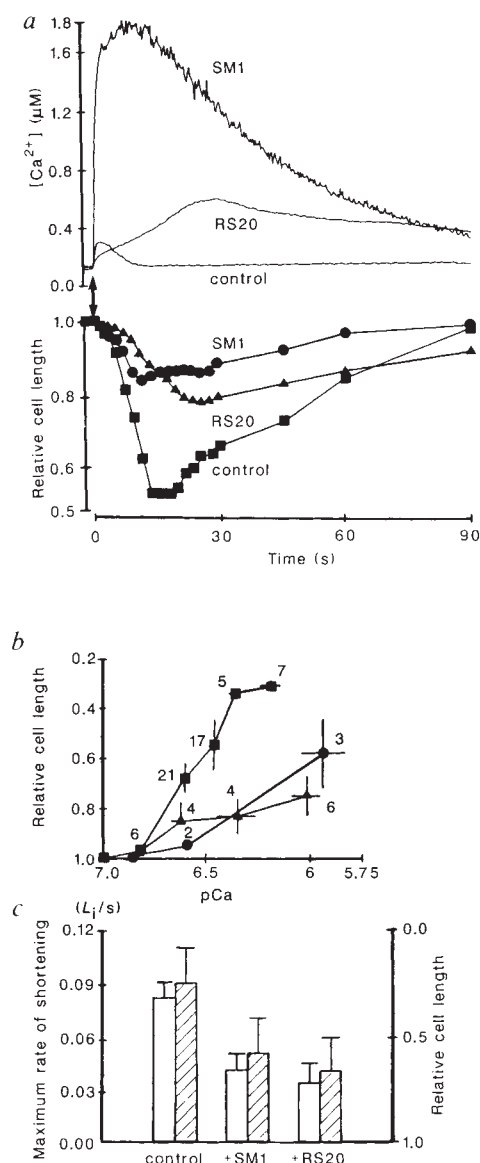


FIG. 4 Effect of SM1 and RS20 inhibitory peptides on  $[Ca^{2+}]$  and contractility changes in response to high  $K^+$  stimulation. *a*, Effect of high  $K^+$  stimulation on  $[Ca^{2+}]$  and cell length in two cells. The control and RS20 responses were obtained from the same cell after impalement by application of  $K^+$ -substituted APS before and then 5 min after injection of peptide. The SM1 results were obtained in another cell 5 min after microinjection of this peptide. Both peptides were injected at a concentration of 5 mM at 40 psi for 3 s and should thus be present at an intracellular concentration of 35  $\mu$ M, given that under these conditions the injected volume was typically 0.7% of the cell. The  $Ca^{2+}$  changes and contractile responses to high  $K^+$  stimulation were quite reproducible: the maximum change in length and  $[Ca^{2+}]$  following injection of MIB was, respectively,  $95.8 \pm 7.5\%$  and  $105.8 \pm 4.5\%$  of that seen before injection ( $n=3$ ).  $K^+$ -substituted APS was applied for 1 s beginning at the point indicated by the vertical arrow. *b*, Effect of RS20 and SM1 peptides on the relationship between peak contractile response and peak  $[Ca^{2+}]$  change. The data have been grouped according to the  $[Ca^{2+}]$  change achieved. Symbols: control (■), SM1 (●) and RS20 (▲). Numbers indicate the number of observations for each group. Note that RS20 and SM1 had no detectable effect on the  $Ca^{2+}$  sensitivity or spectral properties of fura 2 *in vitro*. *c*, Effect of RS20 and SM1 peptides on the contractile responses to  $K^+$  depolarization. The results show the mean values for the maximum extent (open bars) and maximum rate of shortening (cross-hatched bars) for control ( $n=28$ ), SM1-injected ( $n=9$ ) and RS20-injected ( $n=12$ ) cells. Data were included only from cells in which the  $[Ca^{2+}]$  exceeded 500 nM following  $K^+$  depolarization. This selection criterion was applied to ensure that activation of the contractile system was the same for all control cells, because isometric force production is maximal in these cells beyond a  $[Ca^{2+}]$  of 500 nM (ref. 8).

Received 15 November 1988; accepted 16 January 1989.

- Somlyo, A. P. *Circulation Res.* **57**, 497–507 (1985).
- Adelstein, R. S. & Eisenberg, E. A. *Rev. Biochem.* **44**, 921–956 (1980).
- Kamm, K. E. & Stull, J. T. A. *Rev. Pharmacol. Toxicol.* **25**, 593–620 (1985).
- Walsh, M. P. in *Regulation and Contraction of Smooth Muscle* (eds Siegman, M. J., Somlyo, A. P. & Stephens, N. L.) 119–141 (Liss, New York, 1987).
- Dillon, P. F., Aksoy, M. O., Driska, S. P. & Murphy, R. A. *Science* **211**, 495–497 (1981).
- Siegman, M. J., Butler, T. M., Mooers, S. U. & Michalek, A. *Pflügers Arch. ges. Physiol.* **401**, 385–390 (1984).
- Fay, F. S., Hoffman, R., LeClaire, S. & Merriam, P. *Meth. Enzym.* **85**, 284–292 (1982).
- Yagi, S., Becker, P. L. & Fay, F. S. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4109–4113 (1988).
- Williams, D. A., Fogarty, K. E., Tsien, R. Y. & Fay, F. S. *Nature* **318**, 558–561 (1986).
- Williams, D. A., Becker, P. L. & Fay, F. S. *Science* **235**, 1644–1648 (1987).
- Becker, P. L. & Fay, F. S. *Am. J. Physiol.* **253**, C613–C618 (1987).
- Ikebe, M., Stepinska, M., Kemp, B. E., Means, A. R. & Hartshorne, D. J. *J. biol. Chem.* **262**, 13828–13834 (1987).
- Pearson, R. B., Wittenhall, R. E. H., Means, A. R., Hartshorne, D. J. & Kemp, B. E. *Science* **241**, 970–973 (1988).
- Kemp, B. E., Pearson, R. B., Guerico Jr, V., Baguchi, I. C. & Means, A. R. *J. biol. Chem.* **262**, 2542–2548 (1987).
- Lukas, T. J., Burgess, W. H., Prendergast, F. G., Lau, W. & Watterson, D. M. *Biochemistry* **25**, 1458–1464 (1986).
- Ikebe, M., Kemp, B. E. & Hartshorne, D. J. *Biophys. J.* **53**, 182a (1988).
- Fay, F. S. & Delise, C. M. *Proc. natn. Acad. Sci. U.S.A.* **70**, 641–645 (1973).
- Gryniewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440–3450 (1985).
- Kargacin, G. J. & Fay, F. S. *J. gen. Physiol.* **90**, 49–73 (1987).
- Ikebe, M. & Hartshorne, D. J. *J. biol. Chem.* **260**, 13146–13153 (1985).
- Ikebe, M. & Hartshorne, D. J. *Biochemistry* **24**, 2380–2386 (1985).
- Walsh, M. P., Hinkins, S., Dabrowska, R. & Hartshorne, D. J. *Meth. Enzym.* **99**, 279–288 (1988).
- Fay, F. S. & Singer, J. J. *Am. J. Physiol.* **232**, C144–C154 (1977).

ACKNOWLEDGEMENTS. We thank Dr P. L. Becker, K. E. Fogarty and S. Abramson for development of essential software, Dr H. Yamaguchi for guidance in designing optimum electrodes for cell injections, and R. McDermott for secretarial assistance. We also thank Prof. Hiroshi Kuriyama (Kyushu University, Fukuoka, Japan) for giving T.J. the opportunity for postdoctoral work. This research was supported by grants from the NIH (F.S.F., M.J., G.J.K. and D.J.H.), the Muscular Dystrophy Association of America (F.S.F.), the American Heart Association and Syntex (M.I.), National Heart Foundation of Australia (B.E.K.) and NH & MRC (B.E.K.).

## Three-dimensional architecture of the calcium channel/foot structure of sarcoplasmic reticulum

Terence Wagenknecht\*, Robert Grassucci\*, Joachim Frank\*, Akitsugu Saito†, Makoto Inui† & Sidney Fleischer†

\* Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, PO Box 509, Albany, New York 12201-0509, USA  
† Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235, USA

THE calcium channel responsible for the release of  $Ca^{2+}$  from the sarcoplasmic reticulum of skeletal muscle during excitation-contraction coupling has recently been identified and purified<sup>1–4,21</sup>. The isolated calcium channel has been identified morphologically with the 'foot' structures<sup>1,2</sup> which are associated with the junctional face membrane of the terminal cisternae of sarcoplasmic reticulum. *In situ*, the foot structure extends across the gap of the triad junction from the terminal cisternae of the reticulum to the transverse tubule<sup>5</sup>. We describe here the three-dimensional architecture (3.7 nm resolution) of the calcium channel/foot structure from fast-twitch rabbit skeletal muscle, which we determined from electron micrographs of isolated, non-crystalline structures that had been tilted in the electron microscope. The reconstruction reveals two different faces and an internal structure in which stain accumulates at several interconnected locations, which could empty into the junctional gap of the triad junction. The detailed architecture of the channel complex is relevant to understanding both the physical path followed by calcium ions during excitation-contraction coupling and the association of the terminal cisternae and the transverse tubules in the triad junction.

We propose the term junctional channel complex (JCC) for the foot structure, which serves both as the junctional connection



that senses the depolarization signal from the transverse tubule, and as the calcium channel itself. Our method of three-dimensional reconstruction takes advantage of the preferential orientation assumed by the JCC when adsorbed to the carbon support surface of an electron microscope grid<sup>6</sup>. Typically, most of the complexes, as viewed by negative staining, are square shaped and have a characteristic morphology consisting of a dense (strongly stain-excluding) central core region with an apparent 2-nm hole in the centre (Fig. 1*b*). Surrounding the core region are four outer regions (refs 3, 6, 7; see also Fig. 1*b*). To determine the three-dimensional (3D) architecture of the JCC, it is necessary to record each specimen field twice: first, with the specimen grid tilted (by 55° in this study), and second, with the grid untilted (Fig. 1*a, b* respectively). The images of the individual JCC in the tilted micrographs, considered as a set, effectively form a conical-tilt series of the molecule. The method of 3D reconstruction used here has been detailed recently<sup>8,9</sup>.

Analysis of the images obtained for the untilted specimen is necessary for 3D reconstruction. Averaged images (Fig. 1*c*) have enhanced structural details, and when several such averages are determined over independent sets of particles, then the degree of structural preservation can be estimated<sup>10</sup>. Maps of the statistical variance (Fig. 1*d*) of density associated with each picture element of the averages are useful for identifying localized regions of high or low variability among the molecules. More detailed information concerning inter-image differences can be obtained by multivariate statistical analysis and classification techniques applied to the individual images<sup>11,12</sup>. The results of such studies can be summarized as follows. The reproducible resolution of the averaged projections (Fig. 1*c*) has improved relative to our previous study<sup>6</sup> and extends to at least 2.5 nm, which is close to the limits of the negative-staining technique. Localized regions of poorer reproducibility (higher variance) are located in the outermost regions of the averaged images (Fig. 1*d*). The strong fourfold symmetry is apparent in the averages. Multivariate analysis indicated that the JCC images were fairly homogeneous (the set of images could not be repro-

ducibly divided into morphologically distinct subsets). Surprisingly, we found that virtually all of the JCC adsorb to the grid by one of the two faces; the individual images have a definite handedness (recognized by the pinwheel appearance), the sense of which depends on the face of the molecule that has adsorbed. Clearly, the two faces of the JCC which associate with the junctional face membrane of the terminal cisternae of sarcoplasmic reticulum and the transverse tubule are distinct from one another.

Computed representations of the surface topography determined from one of the 3D reconstructions after low-pass filtration to 3.7 nm, the limiting resolution, are shown in Fig. 2*a–c*. To visualize the internal-staining features of the JCC, the reconstruction was also sliced into two halves along a plane perpendicular to the fourfold symmetry axis (Fig. 2*d, e*). In its overall dimensions (27 × 27 × 14 nm) the 3D model shown agrees well with previous measurements made directly from electron-micrograph images<sup>3,6</sup>. The maximal dimension of the reconstruction when viewed normal to the fourfold axis (Fig. 2*c*) is 14 nm. A comparison of this dimension with the value of 12 nm (for the portion extending out from the junctional face membrane) determined from side views of individual foot structures<sup>14</sup>, indicates that the structures have not been severely flattened. It should be noted that, as with all 3D reconstructions of negatively stained specimens, the detailed topography of surface representations depends on the density threshold level chosen to represent the surface. We therefore provide all of the information contained in the reconstruction by means of the sequential sections shown in Fig. 3. Analysis of frozen hydrated and unstained specimens should provide further structural information and allow possible distortions attributable to the negative staining method to be assessed.

The model of the JCC in Fig. 2 represents a volume of 3,800 nm<sup>3</sup>, which, assuming a protein density of 1.37 g cm<sup>-3</sup> corresponds to a relative molecular mass ( $M_r$ ) of 3.1 million. The apparent volume of the reconstructed foot structure depends strongly on the density threshold selected. Independent mass measurements using scanning transmission electron microscopy

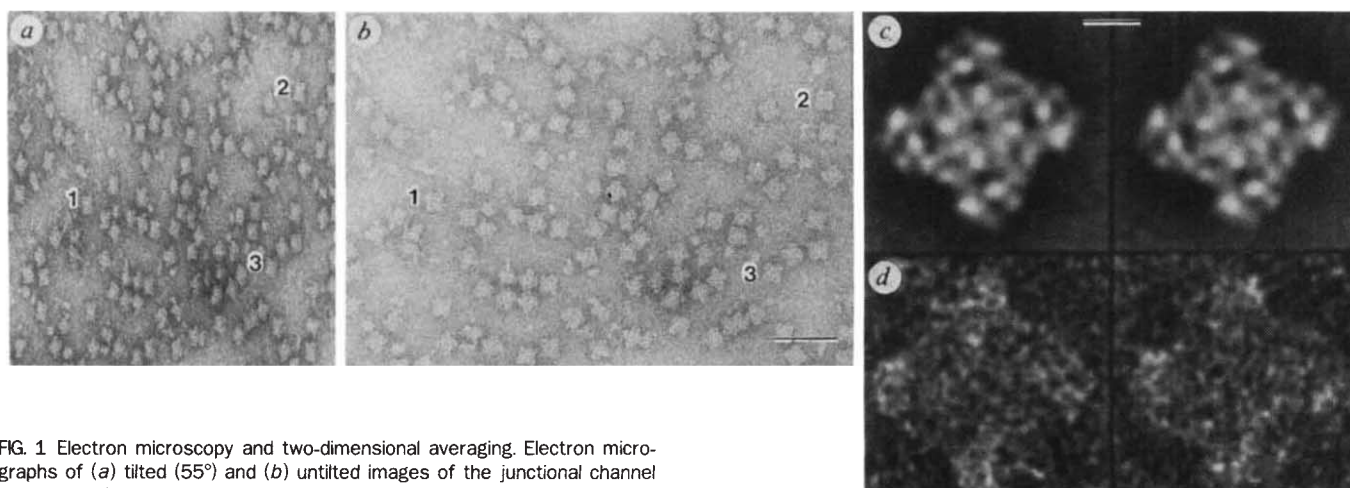


FIG. 1 Electron microscopy and two-dimensional averaging. Electron micrographs of (a) tilted (55°) and (b) untilted images of the junctional channel complexes. Scale bar, 100 nm. Numbers '1', '2', '3' are points of reference. (c), Averaged images obtained for independent sets of untilted channel complexes ( $n=140$  for each, fourfold symmetry not imposed). Scale bar, 10 nm. (d), Variance images corresponding to the averaged images shown in (c). White corresponds to higher variance.

**METHODS.** The JCC from rabbit fast twitch skeletal muscle was purified<sup>1,2</sup> from junctional terminal cisternae SR vesicle<sup>14</sup> with some modifications<sup>16</sup>. The purified JCC was stored frozen at -80 °C in a buffer containing 5 mM potassium phosphate, pH 7.4, 0.5 M KCl, 2 mM dithiothreitol, 5 mg ml<sup>-1</sup> CHAPS and 0.5 µg ml<sup>-1</sup> leupeptin. Negative staining (0.5% uranyl acetate) was by a procedure which sandwiches the specimen between two thin carbon films thereby producing homogenous staining from the top to bottom surfaces of the molecules<sup>17</sup>. In our previous study the more conventional single-carbon layer method was used, and this probably accounts for the

different patterns of variance obtained in the two studies (high variance in the central regions of the images in ref. 6 compared with higher variance in the outer regions found in the present study). A Philips EM420T transmission electron microscope equipped with a low-dose kit and goniometer stage was used. Each field was photographed twice under minimal dose conditions (about 600–800 electrons nm<sup>-2</sup>): first with the specimen grid tilted by 55° and second, with the grid untilted. The electron-optical magnification was ×36,000. For image processing the micrographs were digitized on a Perkin Elmer PDS 1010A flatbed scanning microdensitometer with a 25 µm square scanning aperture (corresponding to 0.69 nm on the specimen). Procedures used for the selection, translational and rotational alignment, and averaging of the JCC images have been described previously<sup>11,12,18,19</sup>.

yield a  $M_r$  of  $2.3 \pm 0.3$  million<sup>15</sup>. This value together with the fourfold symmetry of the channel complex, would be compatible with four polypeptide protomers comprising the JCC and a protomer  $M_r$  of 500,000. A relative molecular mass for the protomer can be estimated from the high-affinity ryanodine binding of the purified receptor to be 384,000 (ref. 2) and 500,000 (ref. 3) based on the reported maximal binding of 650 (ref. 2) and 500 (ref. 3) pmol per mg protein, respectively, and assuming one high affinity binding event per tetramer<sup>2</sup>. The JCC is an oligomer apparently consisting of a single high relative molecular mass polypeptide ( $M_r = 360,000$ –450,000) according to analysis by SDS-PAGE<sup>2-4</sup>.

A square basal platform (14 nm edge by 4 nm thick) projects from one of the fourfold faces of the reconstruction (labelled 'BP' in Figs 2, 3). A similar feature can be inferred from micrographs of negatively stained JCC lying on its side<sup>6</sup> and in

freeze-dried, rotary shadowed specimens<sup>13</sup>. It was previously interpreted as forming the membrane-spanning portion of the JCC<sup>6</sup>. The central stain-filled channel of the reconstruction does not seem to project through the basal platform (Figs 2e and 3), possibly because the JCC is in the closed state and/or because the putative ion channel is too small to be resolved at the resolution attained. An electron microscope study of the junctional feet, *in situ*, by the freeze-fracture technique also provides evidence that the platform-bearing face of the JCC is the side-facing membrane of the sarcoplasmic reticulum<sup>13</sup>. The hydrophobic nature of the platform, presumed to be attributable to the transmembrane region inserted into the junctional face membrane of terminal cisternae of SR, accounts for the preference of this side of the JCC to adhere to the hydrophobic carbon film of the microscope grid.

In the surface representations shown in Fig. 2, the central

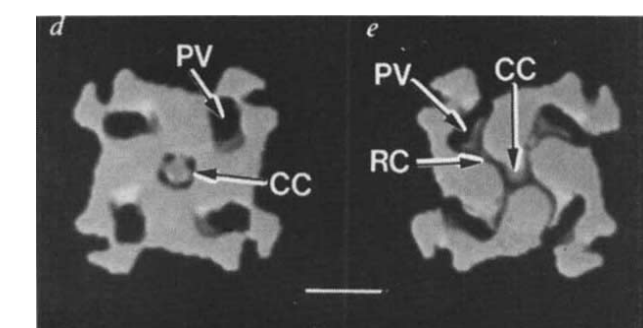
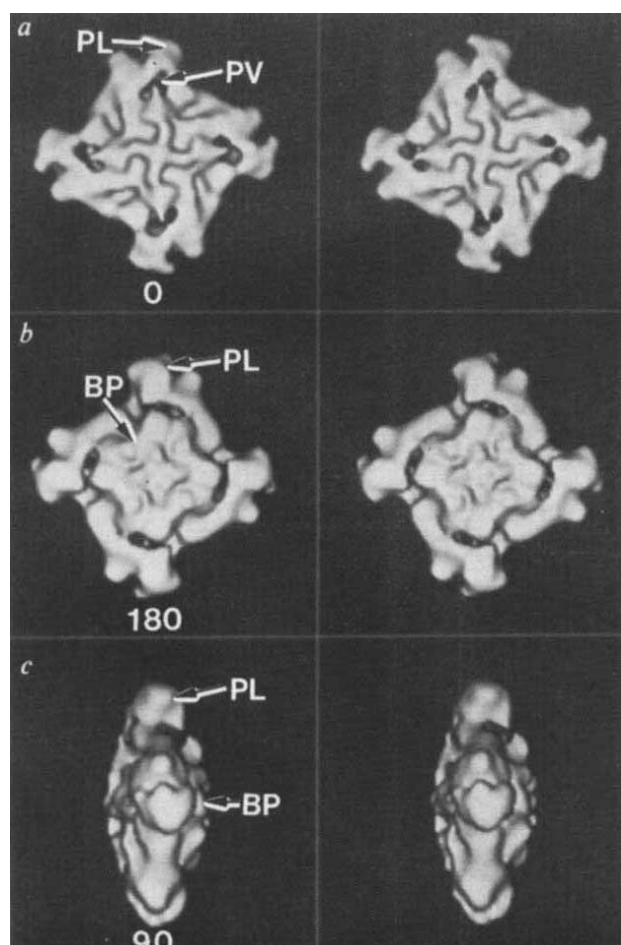
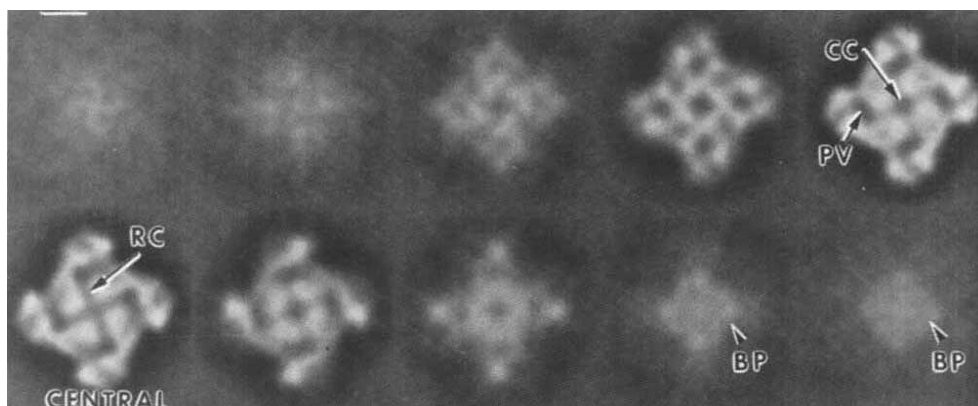


FIG. 2 Computer-generated surface representations of the three-dimensionally reconstructed junctional channel complex. *a*–*c*, Stereo pairs of the reconstruction in various orientations related by rotation about a vertical axis. *d*, *e*, The two complementary halves of the reconstruction after slicing it in half to reveal internal structural features—in (*e*) the view is from the interior of the channel towards the surface which adsorbed to the grid (the 'platform' side) and (*d*) is towards the t-tubule surface. Abbreviations: BP, base platform; PL, peripheral lobes; PV, peripheral vestibules; CC, central channel; RC, radial channels. Scale bar, 10 nm.

**METHODS.** Procedures for 3D reconstruction and computation of surfaces have been described<sup>8,9,20</sup>. The reconstruction shown was determined from 280 particles and has been fourfold rotationally averaged to improve the resolution and signal-to-noise ratio. The resolution of the reconstruction in the X–Y dimensions (parallel to the plane of the electron microscope grid) was estimated to be between 3.5 and 4.5 nm by Fourier ring correlation and phase-residual methods<sup>8,10</sup>. The reconstruction was low-pass filtered to a limiting resolution of 3.7 nm here and in Fig. 3. The surfaces in *d* and *e* were determined for a threshold density 2% higher than was used for *a*–*c*.

FIG. 3 Sections through the three-dimensionally reconstructed volume. Transverse sections (parallel to specimen grid) at 1.38 nm intervals with the central section indicated. The surface formed by the basal platform (BP) adsorbs to the primary carbon-supporting film of the specimen grid. Abbreviations as in Fig. 2 legend. Scale bar, 10 nm.





channel ('CC') does not seem to extend to the surface of either face of the JCC (as mentioned above, this could be a resolution limitation). Branching off from the central channel, near the midplane of the reconstruction, are four radially extending channels (labelled 'RC' in Figs 2e, 3a). These radial channels extend to the four stain-filled vestibules present in the peripheral regions of the JCC (labelled 'PV' in Figs 2 and 3). The peripheral vestibules seem to extend to the surface of the JCC in the 3D reconstruction.

The existence of the radial channels connecting the central channel to the outer vestibules suggests a possible pathway for calcium ion transport when the JCC is in its open state. We assume that there are one or more membrane-spanning channels in the platform of the JCC and that these lead to the central channel. On entering the central channel, ions could flow from the radial channels to the outer vestibules and into the gap regions of the triad junction which is contiguous with the myoplasm. This is reasonable as the other square face of the JCC is associated with the transverse tubule. The model of JCC function should be better resolved when 3D reconstructions of the channel are obtained in open as well as closed conformational states. □

Received 15 November 1988; accepted 25 January 1989.

1. Inui, M., Saito, A. & Fleischer, S. *J. Biol. Chem.* **262**, 1740-1747 (1987).
2. Inui, M., Saito, A. & Fleischer, S. *J. Biol. Chem.* **262**, 15637-15642 (1987).
3. Lai, F. A., Erickson, H. P., Rouseau, E., Liu, Q. V. & Meissner, G. *Nature* **331**, 315-319 (1988).
4. Imagawa, T., Smith, J. S., Coronado, R. & Campbell, K. P. *J. Biol. Chem.* **262**, 16636-16643 (1987).
5. Franzini-Armstrong, C. & Nunzi, C. *J. Muscle Res. cell. Motil.* **4**, 233-252 (1983).
6. Saito, A., Inui, M., Rademacher, M., Frank, J. & Fleischer, S. *J. Cell Biol.* **107**, 211-219 (1988).
7. Saito, A., Inui, M. & Fleischer, S. *Biophys. J.* **53**, 605a (1988).
8. Rademacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. *J. Microsc.* **146**, 113-136 (1987).
9. Frank, J., Verschoor, A., Wagenknecht, T., Rademacher, M. & Carazo, J. M. *Trends biochem. Sci.* **13**, 123-127 (1988).
10. Frank, J., Verschoor, A. & Boublik, M. *Science* **214**, 1353-1355 (1981).
11. van Heel, M. & Frank, J. *J. Ultramicrosc.* **6**, 187-194 (1981).
12. Frank, J., Breaudiere, J. P., Carazo, J. M., Verschoor, A. & Wagenknecht, T. *J. Microsc.* **150**, 99-115 (1987).
13. Block, F. A., Imagawa, T., Campbell, K. P. & Franzini-Armstrong, C. *J. Cell Biol.* **107**, 2587-2600 (1988).
14. Saito, A., Seiler, S., Chu, A. & Fleischer, S. *J. Cell Biol.* **99**, 875-885 (1984).
15. Saito, A., Inui, M., Wall, J. S. & Fleischer, S. *Biophys. J.* **55**, 206a (1989).
16. Inui, M. & Fleischer, S. *Meth. Enzym.* **157**, 490-505 (1988).
17. Stöffler, G. & Stöffler-Meilicke, M. In *Modern Methods in Protein Chemistry* (ed. Tschesche, H.) 409-457 (de Gruyter, New York, 1984).
18. Frank, J., Goldfarb, W., Eisenberg, D. & Baker, T. S. *Ultramicroscopy* **3**, 283-290 (1978).
19. Frank, J., Verschoor, A. & Wagenknecht, T. In *New Methodologies in Studies of Protein Configuration* (ed. Wu, T. T.) 36-89 (Van Nostrand Reinhold, New York, 1985).
20. Rademacher, M. & Frank, J. *J. Microsc.* **136**, 77-85 (1984).
21. Hymel, L., Inui, M., Fleischer, S. & Schindler, H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 441-445 (1988).

ACKNOWLEDGEMENTS. This work was supported by research grants from the National Institutes of Health and the Muscular Dystrophy Association.

## DNA-protein conjugates can enter mitochondria via the protein import pathway

Dietmar Vestweber & Gottfried Schatz

Biozentrum, University of Basel, Department of Biochemistry, CH-4056 Basel, Switzerland

**MITOCHONDRIA** import most of their proteins and small molecules from the cytoplasm<sup>1</sup>. There is some tentative evidence that they import some of their RNAs<sup>1-4</sup>, but it is not known how nucleic acids could enter mitochondria. Here, we show that isolated yeast mitochondria can import a single-stranded or double-stranded 24-base pair piece of DNA whose 5' end is covalently linked to the C-terminus of a mitochondrial precursor protein.

The precursor protein was a fusion protein containing a mitochondrial presequence (derived from the yeast cytochrome oxidase subunit IV precursor) linked to a modified<sup>5</sup> mouse dihydrofolate reductase (DHFR) protein. This modified variant (DV12) can be purified after overexpression in *Escherichia coli*

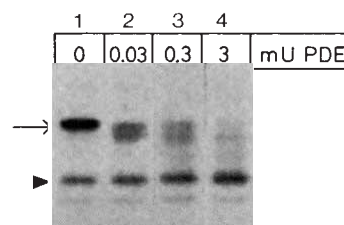


FIG. 1 The oligonucleotide conjugated to the precursor protein can be removed by phosphodiesterase. The mutated DV12 variant<sup>5</sup> of the coxIV-DHFR fusion protein<sup>8,9</sup> coupled to a 24-nucleotide oligonucleotide was digested with the indicated amounts of phosphodiesterase (PDE) and the reaction product analysed by SDS-12% polyacrylamide gel electrophoresis. Arrow: DNA-precursor conjugate; arrowhead: underivatized precursor protein.

**METHODS.** For each digestion, 60 ng of derivatized precursor was incubated for 15 min at 25 °C in 40 µl of 20 mM HEPES pH 7.4, 2.5 mM MgCl<sub>2</sub>, 0.6 M sorbitol with the indicated amounts of *Crotalus durissus* phosphodiesterase (Boehringer, Mannheim). The reaction was stopped with 20 mM EDTA and the sample was analysed by SDS-PAGE followed by fluorography. Coupling procedure: the modified coxIV-DHFR precursor protein containing a unique cysteine as the C-terminal amino acid was purified as described<sup>5</sup> after overexpression and *in vivo*-labelling with <sup>35</sup>SO<sub>4</sub> in *E. coli*. The 24-nucleotide oligodeoxynucleotide 5'-GTACCGAGCTCGAATTGTAATCA-3' was synthesized on a DNA synthesizer (Applied Biosystems, Model 380 B). In the last reaction cycle, the reagent 'aminolink 1' (Applied Biosystems) was conjugated to the 5' end of the oligonucleotide, according to the manufacturer's instructions. In the resulting product, the aminoethyl phosphate linker was conjugated with a phosphoester bond to the 5' hydroxyl group of the oligonucleotide. After deprotection and lyophilization, the modified oligonucleotide was dissolved in water and passed through a Sephadex G10 spin column which had been equilibrated with PBS. The oligonucleotide was incubated in a 1:1 molar ratio with the bifunctional crosslinker maleimidobenzoyl-*N*-hydroxysuccinimide ester for 50 min at room temperature. Unreacted crosslinker was removed by gel filtration on Sephadex G10 (1 ml syringe spin column equilibrated with PBS) followed by a 10 min incubation with 100 mM Tris Cl, pH 7.0 and another gel filtration step. Fifty per cent of the oligonucleotide added initially was recovered in the final eluate. This material was incubated in a 300-fold molar excess, with the purified precursor protein for 25 min at room temperature, followed by a 5 min incubation with 2 mM dithiothreitol to block excess unreacted maleimide groups. Usually 50-80% of the added protein precursor molecules were converted to the DNA-protein conjugate.

and is efficiently imported and cleaved by isolated yeast mitochondria<sup>5</sup>. Its C-terminal unique cysteine was coupled to the 5' end of a 24-nucleotide single-stranded (ss) DNA using a bifunctional crosslinker (Fig. 1). Usually 50-80% of the purified, <sup>35</sup>S-labelled fusion protein could be attached to the single-stranded DNA. As expected, the adduct migrated more slowly in SDS-PAGE gels (Fig. 1, lane 1; Fig. 2, lanes 1 and 2) and its DNA moiety could be removed by phosphodiesterase (Fig. 1).

When the protein-DNA adduct was added to energized isolated yeast mitochondria, about 8% was processed to two smaller forms and became inaccessible to externally added protease (Fig. 2, lanes 3 and 4). An additional 8% became inaccessible without processing (Fig. 2, lane 4). In all cases, inaccessibility to protease was abolished when the mitochondrial membranes were disrupted with Triton X-100 (lane 5). Valinomycin, which de-energizes mitochondria, completely blocked (lane 6). Therefore, 16% of the protein-DNA conjugates entered mitochondria with their protein moiety and 50% of these molecules were processed by the matrix protease. Mitochondria were treated with phosphodiesterase (PDE) to determine whether the oligonucleotide moiety was imported. The conjugates that had been processed by the matrix protease were inaccessible to PDE added externally. All of the unprocessed conjugates however, including the protease-resistant fraction, remained accessible to PDE (Fig. 2, lane 7). Inaccessibility to PDE was abolished by Triton X-100 (lane 9) and was not observed with valinomycin-treated mitochondria (lane 8). We conclude that 8% of the conjugates were processed and completely transported across the mitochondrial membranes and a further 8% were only partly imported such that the protein moiety was inside the mitochon-



dria, and the DNA moiety was outside. The latter conjugates were not cleaved by the matrix protease, probably because the EDTA in the import buffer slows down the matrix protease.

In five independent import experiments with three different precursor preparations and two different oligonucleotide preparations, 5–10% of the added precursor-DNA adducts (2–5 pmol per mg mitochondrial protein) were completely imported and cleaved in the mitochondria. This import efficiency is about five times lower than that of the underivatized precursor (data not shown). At least part of the reduction is caused by residual underivatized oligonucleotide or contaminants from the oligonucleotide preparation as the import efficiency of the DNA-free precursor (which was always present in the import assays, see Fig. 2) dropped by the same extent as that of the DNA-precursor conjugate.

To test whether the processed conjugates that were protected from externally added PDE had reached the soluble matrix space, we subfractionated matrix and membrane proteins of mitochondria by controlled lysis with the non-ionic detergent octyl-polyoxyethylene<sup>6</sup>. We also show the subfractionation of the same precursor protein crosslinked to bovine pancreatic trypsin inhibitor, as a control. This adduct gets stuck in the import channel across both mitochondrial membranes<sup>9</sup>. Figure 3 shows that essentially all the cleaved DNA-protein conjugates cofractionated with the cleaved and fully imported underivatized fusion protein whereas about 95% of the trypsin inhibitor-adduct remained associated with the membranes. Thus, the

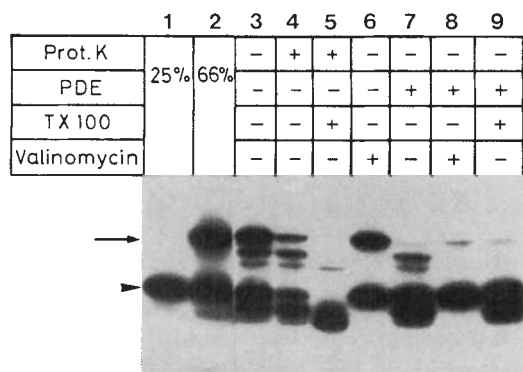


FIG. 2 Single-stranded DNA conjugated to a mitochondrial precursor protein is imported by isolated yeast mitochondria. Lanes 1 and 2: 25% of the amount of underivatized precursor (1) or 66% of the mixture of derivatized and underivatized precursor (2) that was added to the corresponding import reaction. Lane 3, mitochondria that had imported the DNA-protein conjugate; lane 4, as for lane 3 except that mitochondria were treated with 500  $\mu\text{g ml}^{-1}$  proteinase K (Prot. K) for 15 min on ice followed by the addition of 2 mM phenylmethylsulphonyl fluoride before analysis. Lane 5, as for lane 4 but mitochondria were lysed with 0.25% Triton X-100 (TX 100) before protease treatment. Lane 6, mitochondria were incubated with the DNA-precursor conjugate in the presence of the ionophore valinomycin. Lane 7, mitochondria were washed after the import reaction with 0.6 M sorbitol, 20 mM HEPES, pH 7.4, treated with 150  $\text{mU ml}^{-1}$  PDE in 40  $\mu\text{l}$  0.6 M sorbitol, 20 mM HEPES, pH 7.4, 2.5 mM  $\text{MgCl}_2$  for 10 min at 25  $^{\circ}\text{C}$ ; the reaction was stopped by the addition of 25 mM EDTA. Lane 8, as for lane 7 but mitochondria had been incubated with the DNA-precursor conjugate in the presence of valinomycin. Lane 9, same as lane 7 but mitochondria had been lysed with 0.25% of Triton X-100 before treatment with phosphodiesterase. Approximately 50% of the protease-protected DNA conjugates were processed by the matrix protease (lane 4), and protected from externally added phosphodiesterase (lane 7). Arrow: precursor-oligonucleotide conjugate; arrowhead: underivatized precursor. As mentioned earlier<sup>5</sup>, the DV12 fusion protein is cleaved by mitochondria at two sites, yielding two different processed forms.

**METHODS.** The artificial precursor protein DV12 (190 ng) coupled to the oligonucleotide was incubated with 90  $\mu\text{g}$  of isolated energized yeast mitochondria in 200  $\mu\text{l}$  import buffer<sup>10</sup> which was supplemented with 15 mM EDTA to inhibit nucleases in the mitochondrial preparation. Mitochondria were re-isolated by centrifugation and analysed by SDS-12% polyacrylamide gel electrophoresis followed by fluorography.

precursor protein can 'pull' a chain of 24 nucleotides completely across the mitochondrial membranes into the matrix space.

To exclude the possibility that the oligonucleotide could enter mitochondria without the precursor protein, the oligonucleotide was end-labelled with  $^{32}\text{P}$  and incubated with mitochondria under import conditions in the presence and absence of unlabeled, underivatized precursor protein. The labelled oligonucleotide was also incubated without mitochondria under otherwise identical conditions; mitochondria were then added at 0  $^{\circ}\text{C}$  and immediately re-isolated and washed in the same way as the mitochondria from the 'import samples'. In all these cases, only 0.1% of the labelled oligonucleotide was recovered with the mitochondria (data not shown).

To determine whether mitochondria can import double-stranded (ds) DNA, we annealed the DNA moiety of the ss DNA-precursor adduct to a complementary 24-mer oligodeoxyribonucleotide. The annealed product had, reproducibly, a slightly lower mobility in SDS-polyacrylamide gels (Fig. 4a, lanes 1, 2 and top bands of lanes 13 and 14). Three pieces of evidence showed that this mobility shift was due to the ds DNA. Firstly, there was no shift in mobility if the annealing reaction used a non-complementary oligonucleotide of similar length. Secondly, the shift was also abolished if electrophoretic analysis was performed at 55  $^{\circ}\text{C}$  (data not shown). Finally, the ds DNA precursor adduct was digested by exonuclease III, which only cleaves ds DNA, but the ss adduct was not (Fig. 4b).

The ds DNA-precursor adduct was imported by isolated yeast mitochondria as efficiently as the corresponding ss DNA adduct (Fig. 4a). It was cleaved to two smaller forms (lanes 5, 14) which were at least partly inaccessible to externally added protease (lanes 6, 15) or PDE (lane 10) unless detergent was also added (lanes 9, 12); import was abolished by uncoupling the mitochondria (lanes 7, 8, 11). In six independent import experiments, 5–10% of the added ds DNA-precursor adducts were processed and imported as was the case for the ss adduct. The protease-inaccessible unprocessed fraction of the ds DNA-precursor adduct was also resistant to external PDE; however, as Triton X-100 did not alter this resistance to PDE (perhaps because of aggregation of the molecules) the location of the unprocessed, PDE resistant molecules is uncertain.

The slight mobility difference between the ds and ss DNA-precursor adducts was also observed with the corresponding

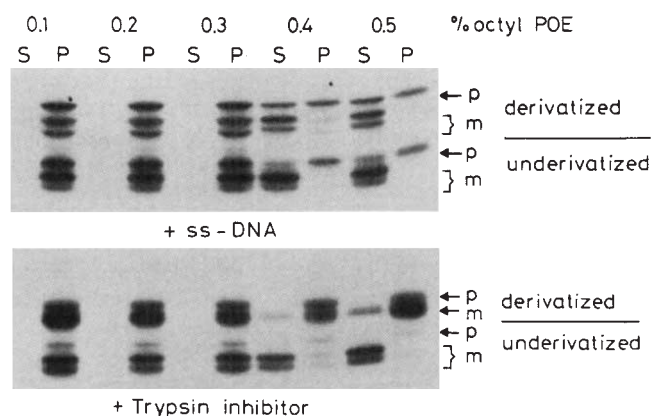


FIG. 3 The ss DNA-conjugate is imported into the soluble matrix space. Mitochondria were allowed to import the DV12 precursor protein which had been conjugated either to the ss DNA molecule (+ss DNA) or to bovine pancreatic trypsin inhibitor (+ trypsin inhibitor). The latter construct became irreversibly stuck across the mitochondrial membranes<sup>6</sup>. Mitochondria were re-isolated by centrifugation and lysed with the indicated concentrations of the non-ionic detergent octyl-polyoxyethylene (octyl POE) as described<sup>6</sup>. After ultracentrifugation, supernatants (S) and pellets (P) were analysed by SDS-12% polyacrylamide gel electrophoresis and fluorography for the distribution of the processed 'mature' form (m) and the uncleaved precursor form (p). As an internal control, import and distribution of the underivatized precursor protein is also shown.

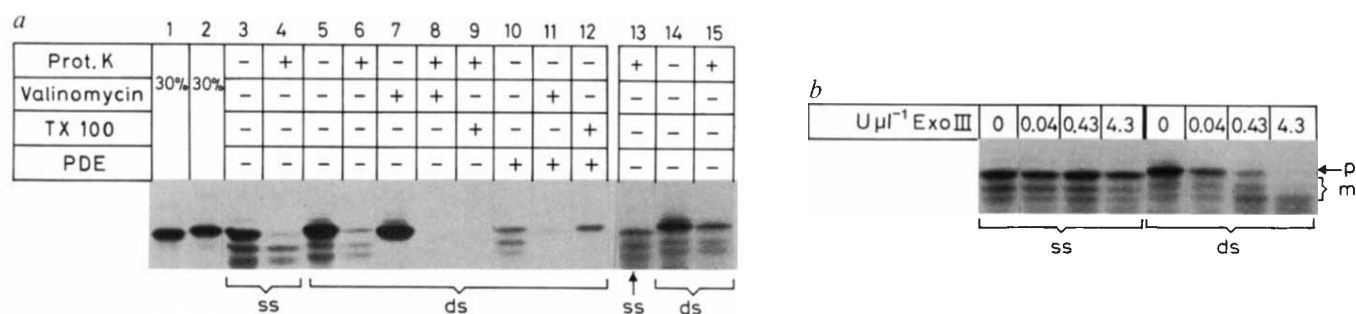


FIG. 4 Double-stranded DNA linked to the mitochondrial precursor protein is imported into mitochondria. **a**, Lanes 1 and 2, 30% of the amount of ss and the ds DNA-protein conjugate, respectively, that was added to the corresponding import reactions. Note that the ds-DNA conjugate (lane 2) migrates more slowly than the ss-DNA conjugate (lane 1). Lanes 3 and 5, mitochondria that had imported the ss- and ds-DNA precursor adducts respectively. Lane 4 and 6, as for lanes 3 and 5, but mitochondria were first treated with proteinase K as described in Fig. 2. Lane 7, mitochondria incubated with the ds-DNA-precursor adduct in the presence of valinomycin. Lane 8, as in lane 7 but mitochondria had been treated with proteinase K before analysis. Lane 9, same as lane 5 but mitochondria were treated with proteinase K in the presence of Triton X-100. Lane 10, mitochondria were treated with phosphodiesterase after import of the ds-DNA-precursor conjugate as described in Fig. 2. Lane 11, as in lane 10 but mitochondria had been incubated with the ds-DNA-precursor adduct in the presence of valinomycin. Lane 12, as in lane 10 but phosphodiesterase treatment was in the presence of Triton X-100. Lanes 13–15 show the results of a separate experiment in which the products were analysed on an 11% polyacrylamide-SDS gel to enhance the mobility difference between ss- and ds-DNA-precursor adducts. Lane 13,

import of the ss-DNA-precursor conjugate, mitochondria treated with protease; lanes 14 and 15, import of the ds-DNA-conjugate without and with protease-treatment of mitochondria, respectively. The portion of the gel displaying underivatized precursor molecules is not shown. **b**, Exonuclease III distinguishes between imported ss and ds DNA-protein conjugates. Only the precursor (p) and mature (m) forms containing ds-DNA were significantly sensitive to exonuclease III.

**METHODS.** Isolated, energized yeast mitochondria (100  $\mu$ g) were incubated under import conditions with either 130 ng of the ss DNA-precursor conjugate (ss) or its ds homologue. The products were analysed as described in the Fig. 2 legend. The ds DNA-protein conjugate was produced by adding the complementary oligonucleotide to the ss DNA-protein conjugate in a 1:1 molar with respect to the first oligonucleotide and annealing for 15 min at room temperature. In **b**, mitochondria were allowed to import ss and ds DNA-protein conjugates re-isolated by centrifugation, washed once in 0.6 M sorbitol, 20 mM HEPES, pH 7.4 and incubated in 40  $\mu$ l 0.6 M sorbitol, 20 mM HEPES, pH 7.4, 0.2 mM  $MgCl_2$ , 0.25% Triton X-100 containing the indicated concentrations of exonuclease III (Boehringer) and analysed by SDS-12% polyacrylamide gel electrophoresis and fluorography.

processed and imported forms (Fig. 4a lanes 4, 5) especially if more porous gels were used (lanes 13, 14). Also, the imported ds DNA-precursor adducts were digested by the ds-DNA-specific exonuclease III (upon disruption of the mitochondria with detergent) whereas the imported ss DNA-precursor adducts were not significantly digested (Fig. 4b). Import into mitochondria thus does not dissociate the two complementary DNA strands.

The double-stranded 24-mer would be expected to have a diameter of 2 nm and form two complete helical turns. The mitochondrial protein import machinery is thus not restricted to the transport of fully extended, linear polypeptide chains<sup>5</sup>. How transport of a rather rigid DNA helix is compatible with the maintenance of a membrane potential is unknown.

Our results do not necessarily imply that nucleic acids enter mitochondria as covalent adducts with imported precursor proteins *in vivo*. They do show, however, that such transport is possible, at least for short pieces of DNA. Protein-mediated transport of nucleic acids might also be exploited with other membrane systems. There is currently great interest in the use of oligonucleotides for inhibiting the expression of individual viral RNAs or mRNAs in mammalian cells<sup>7</sup>; such oligonucleotides could perhaps be introduced into the cytoplasm by covalently linking them to mutated bacterial toxins that have lost their toxicity, but not their ability to pass through the eukaryotic plasma membrane. □

## Reconstruction of an enzyme by domain substitution effectively switches substrate specificity

John E. Houghton\*, Gerard A. O'Donovan\* & James R. Wild

Department of Biochemistry and Biophysics, Texas A & M University, College Station, Texas 77843, USA

**THE** polar domains of the two transcarbamoylases, aspartate transcarbamoylase (ATCase) and ornithine transcarbamoylase, (OTCase) from *Escherichia coli* bind the common substrate carbamoyl phosphate and share extensive amino-acid sequence homology<sup>1,2</sup>. The equatorial domains of the two enzymes differ in their substrate specificity (ATCase binds aspartate, OTCase binds ornithine) and have decreased sequence identity. While addressing the conservation of specific protein interactions during the evolution of these enzymes, we were able to switch one of their amino-acid-specific equatorial domains to produce a viable chimaeric enzyme. This was achieved by the *in vitro* fusion of DNA encoding the polar domain of OTCase to DNA encoding the equatorial domain of ATCase. The resulting gene fusion successfully transformed an *argI*-*pyrB* deletion strain of *E. coli* to pyrimidine prototrophy, giving rise to *Pyr*<sup>+</sup> transformants that expressed ATCase but not OTCase activity. The formation of this active chimaeric enzyme shows that by exchanging protein domains between two functionally divergent enzymes we have achieved a switching in substrate specificity.

In *E. coli*, ATCase (EC 2.1.3.2, the first enzyme distinct to pyrimidine biosynthesis<sup>3</sup>) and OTCase (EC 2.1.3.3, enzyme in step six of arginine biosynthesis<sup>4</sup>) catalyse the condensation of the carbamoyl moiety of carbamoyl phosphate with aspartate

Received 14 December 1988; accepted 12 January 1989.

- Attardi, G. & Schatz, G. A. *Rev. Cell Biol.* **4**, 289–333 (1988).
- Chang, D. P. & Clayton, D. A. *Science* **235**, 1178–1184 (1987).
- Doersen, C.-J., Guerrier-Takada, C., Altman, S. & Attardi, G. *J. Biol. Chem.* **260**, 5942–5949 (1985).
- Suyama, Y. *Curr. Genet.* **10**, 411–420 (1986).
- Vestweber, D. & Schatz, G. *J. Cell Biol.* **107**, 2037–2043 (1988).
- Vestweber, D. & Schatz, G. *J. Cell Biol.* **107**, 2045–2049 (1988).
- Sarin, P. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7448–7451 (1988).
- Hurt, E. C., Pesold-Hurt, B. & Schatz, G. *EMBO J.* **3**, 3149–3156 (1984).
- Eilers, M. & Schatz, G. *Nature* **222**, 228–232 (1986).
- Gasser, S. M., Daum, G. & Schatz, G. *J. Biol. Chem.* **257**, 13034–13041 (1982).

**ACKNOWLEDGEMENTS.** We thank Drs Victoria Hines, Marzell Herold and Ulrich Hommel for the oligonucleotide synthesis and Michele Probst for secretarial assistance. This study was supported by the Swiss National Science Foundation and the US Public Health Service, and a long-term postdoctoral fellowship (D.V.) from EMBO.

\* Present addresses: Department of Biology, Yale University, New Haven, Connecticut 06511, USA (J.E.H.) and Department of Biological Sciences, University of North Texas, Denton, Texas 76203, USA (G.A.O.).



and ornithine respectively. Although native ATCase from *E. coli* and some other enteric bacteria is a dodecamer composed of six catalytic and six regulatory subunits ( $c_6r_6$ ), the functional catalytic subunit is a trimeric aggregate ( $c_3$ ) of relative molecular mass ( $M_r$ ) ( $\sim 100$  K; ref. 5) and is similar to that of OTCase which is  $\sim 110$  K (ref. 6). The structure of the *E. coli* ATCase has been solved to a resolution of  $2.9 \text{ \AA}$  (ref. 7), which has confirmed the division of the catalytic protomer into two distinct tertiary structures: the 'polar domain' ( $c_p$ ) and the 'equatorial domain' ( $c_e$ ), encoded respectively by the 5' and 3' regions of the *pyrB* gene<sup>7-9</sup>. In contrast to the apparent regional conservation of sequence and function within the polar domains of ATCase (*pyrB*) and OTCase (*argI*), their equatorial domains show significantly less homology in their primary sequence, which is consistent with the idea that these domains have differentiated to facilitate the binding of a different substrate.

The marked conservation of amino-acid sequence within the polar domains of ATCase and OTCase ( $>35\%$ ) suggested to us that this domain of OTCase might accommodate the substitution of the equatorial domain of ATCase for that of OTCase. We accomplished this substitution by fusing the two genes *argI* and *pyrB* at sites approximating the polar/equatorial domain junction of their respective polypeptides. The structural *pyrB* gene contains a *Bst*EII restriction endonuclease site (Fig. 1a) that effectively bisects the gene at the domain junction and corresponds to Asp 162 of the translated polypeptide sequence (Fig. 1b)<sup>10</sup>. The *argI* structural gene lacks such a convenient restriction site so we modified this sequence using site-specific mutagenesis to introduce a *Bst*EII restriction site at a position comparable to that in *pyrB* (corresponding to Asp 163 of the OTCase sequence)<sup>1,6</sup>. This modification permitted the division and subsequent 'in frame' fusion of the 3' region of the *pyrB* gene with the 5' portion of *argI*, as described in Fig. 2. We reasoned that the resulting plasmid, pAIB205, might encode a functional transcarbamoylating enzyme in which the polar domain, containing the majority (8 out of 12) of the defined active-site residues<sup>7-9</sup>, would be provided by OTCase, and the equatorial domain (which includes the remaining four active-site residues) would be recruited from ATCase (Fig. 1b).

Transformation of the recombinant plasmid pAIB205 into an *argF*, *argI*-*pyrB* deletion strain of *E. coli* K-12, TB-2 (ref. 11), resulted in  $\text{Pyr}^+$  transformants. When these transformants were cured of the plasmid by treatment with acridine orange<sup>12</sup>, both ampicillin resistance ( $\text{Ap}^r$ ) and pyrimidine prototrophy ( $\text{Pyr}^+$ ) were always lost. Transformation of strain TB-2 with purified pAIB205 simultaneously conferred  $\text{Ap}^r$  and the ability to grow in the absence of exogenous uracil, confirming that the pyrimidine prototrophy was mediated by pAIB205. Strain TB-2 of *E. coli* was used because, in addition to carrying a mutation in *argF* (a gene that encodes an isofunctional OTCase and is present in wild-type *E. coli* K-12 strains<sup>13</sup>), it bears a chromosomal *argI*-*pyrB* deletion<sup>11</sup>, and so any possibility of recombinational events with the gene fusion that might produce a wild-type ATCase enzyme is precluded.

The apparent specific activity of the chimaeric enzyme, when measured in cell-free extracts of strain TB-2 (pAIB205), is  $1.8 \times 10^{-2} \text{ nmol min}^{-1} \text{ per mg protein}$  and is dramatically lower than expected for a plasmid-borne gene expressed from an *argI* promoter<sup>14,15</sup>. Wild-type trimeric ATCase, when expressed from its own *pyrB* promoter in extracts of strain TB-2 (pPBC201; ref. 16) gave an apparent specific activity of  $4.8 \text{ nmol min}^{-1} \text{ per mg protein}$ . As the expression of the chimaeric and wild-type genes is comparable in these two strains, the difference in their observed specific activities presumably reflects altered kinetic properties or instability of the chimaeric enzyme (especially as enzyme activity of the chimaeric enzyme was lost when we attempted further to purify the enzyme). Both alterations would be possible for the fusion protein because, even though all the binding site residues may be present, their position and maintenance within the active site are almost certainly changed.

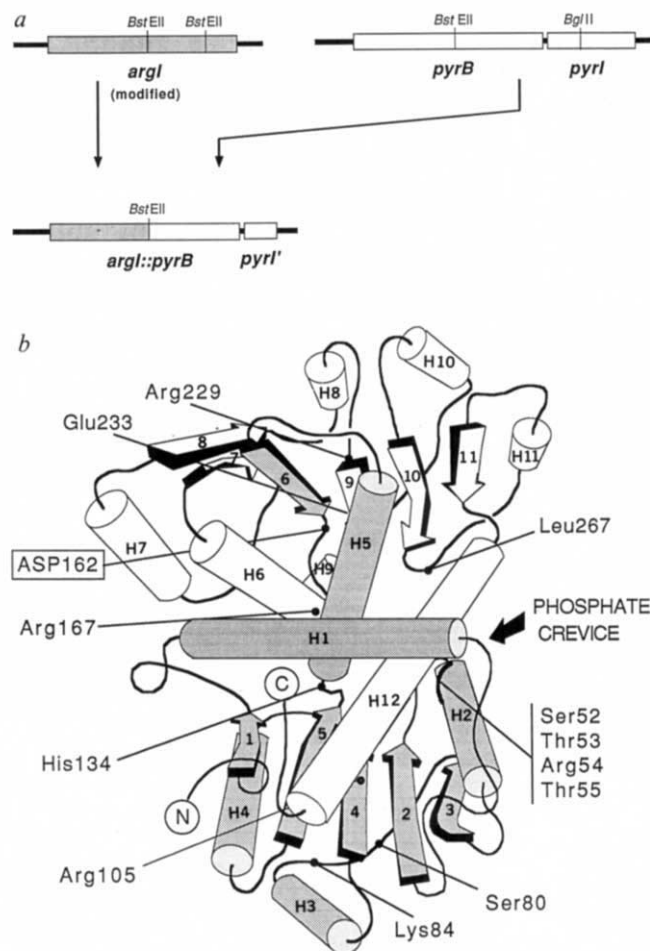
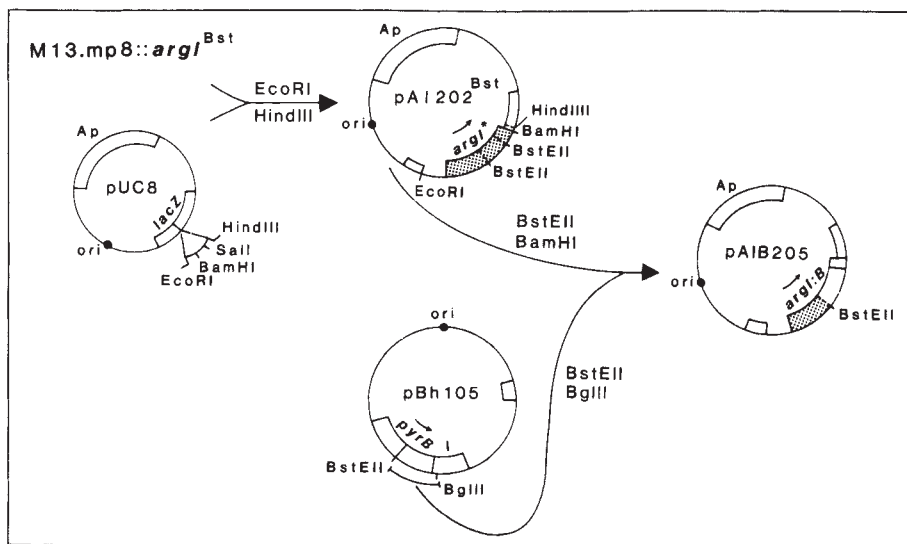


FIG. 1. a, Genetic arrangement of the fusion *argI*::*pyrB* (bottom) as it relates to the parental gene structures of a modified *argI* (top left, shaded) and *pyrB*, *pyrI* operon (top right, clear), showing the location of the *Bst*EII restriction sites used to form the recombinant gene. The *pyrI* gene encodes the regulatory subunit of the ATCase holoenzyme<sup>9,10,16,21</sup> and, although not specifically relevant to the chimaeric construct, contains a unique *Bgl*II restriction site which was used in the formation of the *argI*::*pyrB* fusion, as shown in Fig. 2. The *Bst*EII site on the far left, shown for the modified *argI*, is not found in the wild-type gene and was introduced by site-directed mutagenesis. b, Schematic representation of the structure of a single catalytic polypeptide of ATCase. The chimaeric ATCase is formed from the fusion of the polar domain (shaded structures) of OTCase and the equatorial domain of ATCase (clear structures). The fusion-junction corresponds to Asp 162 of the ATCase sequence. The active site of the enzyme lies within the phosphate crevice and is formed by the asymmetrical insertion of a polar domain from one catalytic subunit into the phosphate crevice of a second subunit. Also shown are the locations of active site residues as defined by X-ray crystallography and other techniques<sup>7-9</sup>.

The recruitment of the polar domain from OTCase, however, has little effect on the  $K_m$  of the chimaeric enzyme for aspartate; at 9 mM aspartate, which is comparable to that of wild-type trimeric ATCase<sup>16</sup>. The  $M_r$  of the active hybrid enzyme is 110 K as determined by Sephadex G-200 column chromatography, which is comparable with the parental enzymes and with other trimeric ATCases<sup>5,16</sup>. Furthermore, the mobility of the catalytically active chimaeric enzyme aggregate on a native polyacrylamide gel is consistent with a trimeric ATCase (Fig. 3, lanes 2 and 3) and differs markedly from a native ATCase holoenzyme (Fig. 3, lane 4). When strain TB-2 was cured of its plasmid pAIB205, no activity was detected (Fig. 3 lane 5). Thus, the ability of this chimaeric gene to encode a fusion protein that aggregates to form a trimer and has aspartate transcarbamoylase activity is strong evidence that the important residues involved



FIG. 2 A schematic representation of the construction of plasmid pAIB205 bearing the recombinant *argI::pyrB* fusion. A *Bst*II restriction site corresponding to the site in *pyrB* was introduced into the phage-borne *argI* gene (M13.mp8::*argI*) by oligonucleotide site-specific mutagenesis. The mutagenic primer that complemented bases 484–502 of the *argI* gene, excepting a two-base mismatch (underlined), <sup>5'</sup>GCAGGTGACCTGCGTAACA<sup>3'</sup>, was synthesized using an Applied Biosystems DNA synthesizer. Introduction of the *Bst*II restriction site into the *argI* gene was achieved by priming a single-stranded M13.mp8::*argI* template containing the nonsense strand of *argI* with the mutagenic primer, and transforming the resulting replicative forms into JM101 (refs 23 and 24), according to the procedure of Zoller and Smith<sup>22</sup>. Transformants were initially screened for the presence of the two-base mutation by selective hybridization of their phage DNA with labelled mutagenic primer<sup>22</sup>. The presence of the mutations was then confirmed by restriction-digest analysis (data not shown). The modified *argI* gene was cloned into the *Eco*RI, *Hind*III restriction sites within the multiple cloning site of plasmid pUC8 (refs 23 and 24), giving rise to plasmid pAI202 which now carried the desired *Bst*II site in the *argI* gene. Subsequently, plasmid pBh105 (ref. 16), bearing the *pyrB* operon, was digested with *Bst*II and *Bgl*II. The resulting 640-base pair *Bst*II/*Bgl*II fragment was purified from a 0.7% w/v agarose gel and ligated into plasmid pAI202 which had been cleaved with *Bst*II and *Bam*HI, giving rise to plasmid pAIB205. This plasmid was transformed into *E. coli* K-12, strain TB-2 (refs. 11), bearing a chromosomal *argI*–*pyrB* deletion. The resulting *Pyr*<sup>+</sup> transformants were all Ap<sup>r</sup> but remained Arg<sup>–</sup>. The



transformed plasmid pAIB205 was subsequently purified over a caesium chloride gradient and the construction verified by restriction enzyme-digest analysis (data not shown). The gene fusion was sequenced in its entirety according to the dideoxy sequencing methods of Sanger<sup>25</sup>, to confirm that the desired in-frame ligation had been achieved, and also to see if any secondary mutations had occurred as a result of selective pressure. Only one non-invasive transition mutation (T→C) was found at position 503 of the fusion DNA sequence.

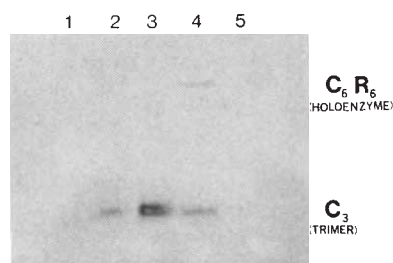


FIG. 3 Similar mobilities of chimaeric and native ATCases on a non-denaturing polyacrylamide electrophoretic gel ATCase activity was identified by inorganic phosphate released after precipitation of orthophosphate to an insoluble lead salt, as described previously<sup>5</sup>. Lane 1, cell-free extract from *E. coli* K-12 strain TB-2 (deleted for *argI* and *pyrB*). Lane 2, cell-free extract from TB-2 transformed with plasmid pAIB205 bearing the chimaeric gene. Lane 3, cell free extract from TB-2 transformed with plasmid pPBC201 bearing the *pyrB* gene, which encodes the catalytic trimer of ATCase<sup>16</sup>. Lane 4, cell-free extract from TB-2 transformed with the plasmids pPBC201 (carrying the *pyrB* gene) and pBr301 (carrying the *pyrI* gene that encodes the regulatory dimer of ATCase<sup>16</sup>). Cells carrying both plasmids produced trimeric (*c*<sub>3</sub>) and holoenzymatic (*c*<sub>6</sub>*r*<sub>6</sub>) aggregates of ATCase *in vivo*. Lane 5, cell-free extracts from TB-2 cured of plasmid pAIB205, which demonstrates that the ATCase activity observed in lane 2 arises from the chimaeric construct borne on the plasmid.

in subunit interactions and in the active site have been conserved throughout the functional differentiation of ATCase and OTCase. Of the 62 residues in ATCase that are conserved in the polar domain of OTCase<sup>1,2</sup>, and are present in the fusion protein, 40 have known structural or purported catalytic functions. Moreover, of the 25 known protein–protein interactions in ATCase, only six have no apparent counterparts in the fusion protein. Interestingly, all contacts that are not conserved in the chimaeric protein aggregate form quaternary interactions (inter-subunit: c1–c3), whereas all the residues that could form tertiary interactions (inter domain: c<sub>p</sub>–c<sub>e</sub>) are apparently maintained. All the known active-site residues in ATCase<sup>7–9</sup> (depicted in Fig. 1b) are present in the gene fusion or have homologous replacements from the polar domain of OTCase<sup>1,2</sup>.

In addition to confirming some important structural and functional relationships between the differentiated ATCase and OTCase enzymes (previously only inferred from sequence homology<sup>1,2,17</sup>) the formation of a chimaeric enzyme by genetic manipulation of the *argI* and *pyrB* genes suggests a number of evolutionary possibilities. The chimaeric enzyme provides an experimental model for the formation of ancestral genes by fusion of distinct genetic modules<sup>18</sup>, and a system for studying the molecular processes involved in cassette recruitment during evolution<sup>19</sup>, which would allow for the potential substitution of subdomain modules to effect changes in substrate specificity<sup>20</sup>. Furthermore, this hybrid provides the genetic template for an active but impaired enzyme whose catalytic efficiency may be optimized by selective pressures, resulting in sequence changes which can be monitored and evaluated accordingly. □

Received 10 October 1988; accepted 23 January 1989.

- Houghton, J. E., Bencini, D. A., O'Donovan, G. A. & Wild, J. R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4864–4868 (1984).
- Van Vliet, F. *et al. Nucleic Acids Res.* **12**, 6277–6299 (1984).
- Gerhart, J. C. & Schachman, H. K. *Biochemistry* **4**, 1054–1062 (1965).
- Legrain, C. *et al. Eur. J. Biochem.* **80**, 401–409 (1984).
- Foltermann, K. F., Beck, D. A. & Wild, J. R. *J. Bact.* **167**, 285–295 (1986).
- Bencini, D. A. *et al. Nucleic Acids Res.* **11**, 8509–8518 (1983).
- Krause, K. L., Volz, K. W. & Lipscomb, W. N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 643–1647 (1985).
- Honzatko, R. B. *et al. J. molec. Biol.* **160**, 219–263 (1982).
- Kantrowitz, E. R. & Lipscomb, W. N. *Science* **241**, 669–674 (1988).
- Hoover, T. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2462–2466 (1983).
- Roof, W. D., Foltermann, K. F. & Wild, J. R. *Molec. Gen. Genet.* **187**, 391–400 (1982).
- Hirota, Y. *Proc. natn. Acad. Sci. U.S.A.* **46**, 57–64 (1960).
- Glansdorff, N., Sand, G. & Verhoef, C. *Mutat. Res.* **4**, 743–751 (1967).
- Cleary, M. L., Garvin, R. T. & James, E. *Molec. Gen. Genet.* **157**, 155–165 (1977).
- Lissens, W., Cunin, R., Kelker, N., Glansdorff, N. & Pierard, A. *J. Bact.* **141**, 55–66 (1980).
- Foltermann, K. F., Shanley, M. S. & Wild, J. R. *J. Bact.* **157**, 891–898 (1984).
- Baur, H., Stalon, V., Falmagne, P., Luethi, E. & Haas, D. *Eur. J. Biochem.* **166**, 111–117 (1987).
- Nyuno, H., Broglie, K. E. & Lusty, C. J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2244–2246 (1985).
- Gilbert, W. *Nature* **271**, 501 (1978).
- Richards, J. H. *Nature* **323**, 187 (1986).
- Paauz, C. D., Karels, M. J., Navre, M. & Schachman, H. K. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4020–4024 (1982).
- Zoller, M. J. & Smith, M. *Meth. Enzym.* **100**, 468–500 (1983).
- Vieira, J. & Messing, J. *Gene* **19**, 259–268 (1982).
- Vieira, J. & Messing, J. *Gene* **19**, 269–275 (1982).
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161–178 (1980).

ACKNOWLEDGEMENTS. We thank Mark Shanley, Karen Kedzie and Tony Forster for help in preparing this manuscript. The research was supported in part by the National Institutes of Health and the Robert A. Welch Foundation.

# Steps up for the student

John Tilley

**Statistical Physics.** By Tony Guénault. *Routledge: 1988. Pp.186. Pbk £5.95, \$17.95.*

**Statistical Physics, 2nd edn.** By F. Mandl. *Wiley: 1988. Pp.385. Hbk £44, \$88; pbk £11.95, \$23.90.*

**Statistical Mechanics.** By B.K. Agarwal and Melvin Eisner. *Wiley Eastern: 1988. Pp.270. £26.95, \$39.95.*

It is surely not just folklore that generations of physics students have found statistical mechanics to be one of the hardest parts of their subject to grasp at first meeting. Some of the difficulties originated from the custom of following the historical sequence by teaching classical statistical mechanics (phase space, Liouville's theorem and so on) before the quantum version. However as Guénault clearly demonstrates in his short text, it is simpler to base an introduction on ideas from quantum mechanics such as energy levels, single-particle quantum states and degeneracy. Teachers will find here a persuasive argument in favour of the quantum approach preceding the classical.

The author's goal is to show fairly quickly what statistical mechanics can do in a wide variety of applications, without dwelling upon its theoretical foundations. The Boltzmann distribution is derived first, then the three ideal-gas distribution functions by the method of grouping single-particle states of almost identical energy. The applications include adiabatic demagnetization, black-body radiation, liquid  $^3\text{He}$  and liquid  $^4\text{He}$ , and ferromagnetism. Students will be attracted by the informal, almost racy style. Anyone who masters this text and copes with the well-chosen illustrative problems will have learned much important physics, and will be equipped to tackle a deeper and more mathematically sophisticated treatment of the field. It seems a pity, therefore, that there is no bibliography to point the reader onward to more advanced texts.

Judging by the introductory textbooks that have been published in the past 20 years or so, there is a growing tendency to treat thermal physics as an entity instead of dealing with thermodynamics and statistical mechanics separately. Mandl's text, first issued in 1971 and now revised, is of that genre. One justification for the unified approach, manifest in this carefully structured account, is that it enables the presentation of statistical mechanics as the bridge between the microscopic quantum view of matter and the macroscopic description used by thermodynamics. It is also an obvious way to show how each topic illuminates the other, examples being the quantum definition of entropy and the link between the partition function and the free energy.

Mandl's book is more demanding of the reader than Guénault's: there is more emphasis on fundamentals and a wider selection of theoretical concepts. For instance, the Gibbs distribution is stressed, and its advantage over the grouping-of-states method in obtaining the Fermi-Dirac and Bose-Einstein distribution functions is clearly shown. The author includes several important applications, some as subjects for the challenging problems at the end of each chapter. The section containing hints for solving

the problems is a valuable feature.

More advanced still, the text by Agarwal and Eisner is aimed at the advanced undergraduate and postgraduate. It is certainly not a book for beginners, although it has the virtue of setting out involved mathematical arguments in full detail. The first chapter is devoted to classical statistical mechanics, and the rest then follows a straightforward exposition of Gibbs's ensemble theory. Problems usefully fill out the material of each chapter. There are brief mentions of topics of current interest, such as the quantum Hall effect and the renormalization group theory of phase transitions. However, this book (which is rather unattractively produced) faces strong opposition in the form of established works such as Huang's *Statistical Mechanics* (Wiley) or *Statistical Physics* in the Landau and Lifshitz series (Pergamon). □

John Tilley is a Lecturer in the Department of Physics, University of Essex, Colchester CO4 3SQ, UK.

## Course variety

R.W. Whitworth

**Fundamentals of Solid State Physics.**

By J. Richard Christman. *Wiley:1988. Pp.518. Hbk £41.55, \$46.85; pbk £13.95, \$12.85.*

**Basic Solid State Chemistry.** By Anthony R. West. *Wiley:1988. Pp.415. Hbk £50, \$97.75; pbk £13.50, \$26.40.*

**Lectures on the Electrical Properties of Materials, 4th edn.** By L. Solymar and D. Walsh. *Oxford University Press:1988. Pp. 465. Hbk £35, \$55; pbk £15, \$27.50.*

**Electrons in Solids: An Introductory Survey, 2nd edn.** By R. H. Bube. *Academic:1988. Pp.315. £27.50, \$39.50.*

**An Introduction to Solid State Diffusion.** By Richard J. Borg and G.J. Dienes. *Academic:1988. Pp.360. £33, \$49.50.*

WE NOW know a lot about the solid state: what aspects of that knowledge should be covered in an undergraduate course?

Many elements and compounds can be fabricated in a variety of solid forms, and have diverse mechanical, thermal, electrical, optical, magnetic and chemical properties. The basic science of these properties is relevant to all materials — from steel to concrete, from plastics to carbon fibres, from micro-electronics to optical components. How then does one construct the syllabus for what is just one component in an undergraduate course in chemistry or in physics? How should one treat the topics that will be needed by students of materials science or electrical engineering who have not had the background courses in physics or chemistry?

J. Richard Christman claims that his book, *Fundamentals of Solid State Physics*, is an appropriate text for a "first course in solid state physics at the advanced undergraduate level". It covers most of what it has been customary to teach physics undergraduates, plus a bit more. The emphasis is therefore on lattice vibrations and the band theory of electrons, preceded by crystal structure and bonding, and leading on to optical and magnetic properties. There is the usual section on superconductivity and an introduction to semiconductor devices, in which the author discusses the p-n junction and touches on some forms of transistor. As is rather common in physics courses, properties which depend on crystal anisotropy and on lattice defects receive only token attention; the page on dislocations is inadequate and misleading.

The book is a genuine contender in the market that has largely been monopolized for a generation by successive editions of Charles Kittel's *Introduction to Solid State Physics*, but I would admire the student who could successfully use it in a first course on the subject. The students I teach tend to need an intermediate step between basic physics and the solid state at the advanced undergraduate level.

A chemist's view is provided by Anthony R. West in *Basic Solid State Chemistry*. As one would expect, West puts more emphasis on structure and bonding, and there is discussion of a much wider range of materials, including the newly discovered high- $T_c$  superconductors. Some crystallography is covered, with simple ideas being developed quite a long way, although the reciprocal lattice is not mentioned. The reader is introduced to phase diagrams, to



many physical properties and to modern techniques such as ESCA, EXAFS and DTA. There are some errors in the coverage of topics which are more often the province of physics. Most seriously, the process described for electrical conduction in metals appears to require thermal activation; another example is that superconductors are given the familiar value of  $-1$  for their magnetic susceptibility, but this is not correct for the old-fashioned system in which susceptibility is defined in the book.

L. Solymar and D. Walsh's book is for students of engineering who need to know about the materials used in devices. The fourth edition of their *Lectures on the Electrical Properties of Materials* includes two new chapters which extend the subject range to include optical as well as electrical devices. The style of writing has the informality hinted at by the title, and students will find the book very readable. Engineering undergraduates encounter devices earlier in their course than do their counterparts in physics, but without the same background. The authors therefore start with a survey of the relevant quantum properties of electrons. They proceed to the electronic properties of materials without getting bogged down in detail, and in my opinion do this rather well.

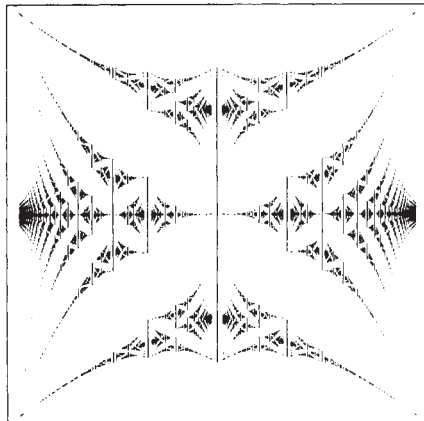
Richard H. Bube's *Electrons in Solids* is intended for students of engineering and also materials science. It is subtitled *An Introductory Survey*, and in about half the number of pages takes the subject from the same starting point as Solymar and Walsh to a deeper level, with parts that are more appropriate to first-year graduates.

The first five chapters provide a survey of the physics of waves, electromagnetism and atomic structure, with quite a lot of mathematical detail. This may be useful for reference or revision, but I doubt whether many people would start their study of these topics from here. The author then deals with the free-electron theory of metals, band theory, and the optical, electrical and magnetic properties of crystals — I recommend the chapter on optical properties as providing a particularly helpful survey of what is sometimes a confusing topic. The reader is also introduced to a number of different semiconductor devices, including modern ones involving multilayer structures or 'band gap engineering'. This second edition differs from the first only in updating, some changes in presentation and the inclusion of a good number of problems and worked examples.

*An Introduction to Solid State Diffusion*, by Richard J. Borg and G.J. Dienes, is really in a different category to the other books reviewed here because it concentrates entirely on one topic, the processes by which atoms migrate through crystalline materials. This is an important topic at the interface of physics and chem-

istry, being relevant to solid-state reactions, corrosion, and the fabrication and degradation of materials and devices. The book could serve as an introduction to the potential specialist or as a simple reference work for the rest of us. A background knowledge of first-year physical chemistry is assumed, but I feel that those who will get most from the book will almost certainly have wider experience.

The authors show up the wealth of material in solid-state science by the variety of different cases that have to be considered. But it seems a pity that although their treatment of semi-



Pattern in physics — an energy diagram for an electron travelling in a weak periodic potential. The diagram appears in A.B. Pippard's *Magnetoresistance in Metals*, published earlier this year by Cambridge University Press.

conductors includes the concentrations of electrons and holes, the diffusion of these charge carriers (which is so beautifully demonstrated in the Haynes-Shockley experiment) is considered to be outside the scope of the book. It is also unfortunate that, in the attempt to maintain generality while avoiding the use of conventional vector analysis, the theory of diffusion has been based on some strange and inadequately explained equations.

It is intriguing, and surely says something about the state of the subject, that here we have an "Introduction" to one aspect of the solid state that is more than half as long as Christman's "Fundamentals", and yet that aspect is not even mentioned by Christman. Which brings us back to the question of how to design the syllabus for an undergraduate course. Students need the breadth that comes from contact with specialists in more than one area, but they are already being swamped with detail. Each of these books provides a credible structure for courses with a particular emphasis, but the search continues for a text that combines balanced breadth with reliability and simplicity. □

R.W. Whitworth is a Senior Lecturer in the School of Physics and Space Research, University of Birmingham, Birmingham B15 2TT, UK.

## Noted approach

D.J. Raine

**General Relativity: A Guide to Its Consequences for Gravity and Cosmology.** By J.L. Martin. Ellis Horwood: 1988. Pp. 176. £32.50. Distributed in the United States by John Wiley, \$63.95.

THERE is, of course, no ideal book for the course. The standard works are too encyclopaedic, the popular expositions lack mathematical detail, the undergraduate texts are too dull or too idiosyncratic. So, over the years, one develops a set of lecture notes which become handouts to appreciative students, who find them just what they need to answer the examination questions. A publisher is found and another textbook enters the crowded market place.

There are, basically, only two ways to expound general relativity. One starts with the machinery of riemannian geometry, building up to Einstein's equations and their solutions. The other begins with the non-relativistic affine geometry of newtonian gravity, adding the metric when relativity is introduced. Either can come with arbitrary amounts of mathematical leavening.

Dr Martin chooses the first approach. In a book addressed principally to physicists, he keeps the mathematics strictly to a minimum. And he gets to the physics as rapidly as possible by dealing with the Schwarzschild metric and the classical tests of the theory before the curvature tensor and field equations. Then there is additional material on black holes, gravitational collapse, gravitational radiation and some cosmology. Nevertheless, as usual in this approach, Einstein's theory does tend to emerge as a dog wagged by a tail, a theory that just happens to satisfy observations.

But there is a problem — lecture notes are not textbooks. It is not so amusing to read that there is no real evidence for the hot big bang, or to find the equation of state of inflationary models dismissed as being improbable. Should not the student who is interested in physics, and who has worked through a course in general relativity, know the importance of the binary pulsar and be able to recognize the contemporary version of the cosmological constant problem?

Let me emphasize that this is not a bad book; on second reading I began to warm to the attention to detail, the care for students' intellectual digestive capacities, the no-nonsense, clear style. But I wonder how many will think it is the ideal book for their course. □

D.J. Raine is a Lecturer in the Department of Astronomy, University of Leicester, Leicester LE1 7RH, UK.

# On the practical side

P.V.E. McClintock

## Experimental Physics: Modern Methods.

By R.A. Dunlap. Oxford University Press: 1988. Pp. 377. \$49. To be published in Britain in April, £45.

PHYSICS is *a fortiori* an experimental subject. Take away the experiments and one is left with mathematics; and the mathematics, however beautiful, may not necessarily relate to the real universe that physicists endeavour to describe. That is why all undergraduate students of physics, even aspiring theorists, are at some stage expected to undertake laboratory experiments.

Organizers of laboratory classes encounter a number of problems. Commonly, because of the expense of providing multiple sets of each experiment, the students in a class will be working on a whole range of different experiments at any moment, rather than all doing the same one. Thus a parallel lecture course is inherently difficult to relate to the laboratory work. Reference to a necessarily wide range of standard texts is useful (and also of some value in familiarizing students with libraries), but this is often inconvenient, and sometimes even unhelpful, because the required text may very well not be available when it is needed.

It would be of immense value, therefore, to be able to recommend students to buy their own copies of a single book containing relevant information about most of the experimental work they will undertake. R.A. Dunlap's *Experimental Physics* is largely intended to be such a book. It owes its origins to a full-year junior-level course in experimental physics at Dalhousie University, Nova Scotia; the level is broadly comparable with a second-year undergraduate course in Britain. The 14 chapters cover a huge range, but concentrate particularly on topics in solid-state physics, optics and nuclear physics. The text is designed to act as a complete introduction for students majoring in physics and also as a "reference work for technicians throughout a professional career".

How far has Professor Dunlap succeeded in these aims? On the plus side, the book provides a clear and friendly exposition of many experimental topics — semiconductors, electronics, signal processing, thermometry, cryogenics and magnetism — some of which have been introduced relatively recently to the undergraduate laboratory. There are helpful references to standard texts relating to each of the chapters, for students who are interested or who require more detailed informa-

tion. Of course, in a book of this kind almost every reader will be able to identify his or her own list of 'missing' topics: for example, there is not much discussion of experiments in thermodynamics or critical phenomena, or of the use of bridges for electrical measurements. The quality of production is mostly good, although a few of the diagrams strike me as crude and certainly fall far below the usual standards one associates with Oxford University Press.

One must seriously question the longevity being claimed for the book. Digging out the equivalent text in use during my own undergraduate days (B.L. Worsnop's and H.T. Flint's admirable *Advanced Practical Physics for Students*, 9th edn,

Methuen, 1957), I was struck by how much of the landscape of experimental physics has been transformed during the intervening period; in fact, that book must already have seemed dated to the point of quaintness a decade or more ago. In this context, to say that *Experimental Physics* will be suitable for use throughout a professional career must surely be construed as publisher's hyperbole. Nonetheless, this first edition of Professor Dunlap's book is much to be welcomed. It should provide valuable reading and support for large numbers of physics students, over the next several years at least. □

P.V.E. McClintock is a Reader in the Department of Physics, University of Lancaster, Lancaster LA1 4YB, UK.

## Simply . . . optics

John Macdonald

**Optics.** By K.D. Möller. University Science Books, 20 Edgehill Road, Mill Valley, California 94941/Oxford University Press, UK:1988. Pp.644. \$48, £35.

**Optics, 2nd edn.** By F.G. Smith and J.H. Thomson. Wiley:1988. Pp.320. Hbk £36.50, \$73; pbk £11.95, \$23.90.

**Optics, 3rd edn.** By W.T. Welford. Oxford University Press: 1988. Pp.155. Hbk £17.50, \$39.95; pbk £8.95, \$14.95.

IF YOU have sweated blood and tears in writing an undergraduate textbook on optics, why agonize over the choice of title? Better surely to follow some illustrious predecessors and simply call it *Optics*. That is what each of the authors of the books reviewed here has done, one (Möller) in a brand new book, the other two in new editions of volumes originally published in the 1970s. In doing so, they must invite comparison with other new editions with exactly the same title and a similar intended audience, such as those by Hecht (Addison-Wesley) and Klein and Furtak (Wiley). And presumably the Pedrotti brothers would wish their new undergraduate textbook to be ranked in the same company, even though they have been rather verbose in their title, *Introduction to Optics* (Prentice-Hall).

Looking at the three books side by side, one is immediately struck by the differences in style and format, which are reflected in their relative prices. So in fairness to the authors, readers of this review should remind themselves of the prices shown above before going further.

Möller is a well-produced, larger-format book with colour plates, and even a rainbow hologram. The others are both part of standard series of physics textbooks (Welford from the Oxford Physics Series, Smith and Thomson from the Manchester Physics Series), and they are

constrained in size, style and lavishness by the necessary uniformity of the series. All three aim to cover a wide syllabus in undergraduate optics, although with a number-of-pages ratio of 1:2:4 (Welford: Smith and Thomson:Möller) the same breadth and depth cannot be maintained in all of them. Welford acknowledges the brevity with which he covers some topics, and suggests that his book might correspond to a first-term university course. The other authors see theirs as supersets of an undergraduate syllabus in optics, and invite the lecturer to select a course from among the topics included.

Hooray! In geometrical optics, all three books employ the cartesian sign convention for distances, which is used by every self-respecting optical designer (whereas, oddly enough, the three competitors mentioned earlier use schoolboy style conventions). All deal with Fermat's principle, but Möller states it only in Fermat's original form of minimum path, rather than as the correct stationary path. This lends force to his use of Feynman's life-guard analogy, which is certainly memorable, but it provides little insight into the physics underlying Fermat's principle. This is no mere pedantry: in many straightforward cases, rays do follow maximum paths.

Möller spends some time dealing with the matrix approach to paraxial ray-tracing, whereas this is not covered by the other two books. I personally find little to commend the matrix approach — in my experience it is rarely used by professionals in geometrical optics — but I acknowledge that it is a popular topic in optics textbooks (for example in Klein and Furtak). The Smith-Helmholtz-Lagrange optical invariant is surely an important item in geometrical optics; I could not find it dealt with by either Möller, or Smith and Thomson. Understandably at this level, none of the authors gives a long account of aberrations, but Welford's is particularly brief.

On the subject of diffraction theory,



all of the books contain a reasonable introduction to Fraunhofer diffraction (although Welford says little on the theory of diffraction gratings), but the depth of coverage of Fresnel diffraction shows greater variation. Welford assigns the topic only two pages, omitting, for example, the Fresnel zone plate and the Cornu spiral. Möller's treatment is the most detailed.

The best account of interference is perhaps provided by Smith and Thomson, although some interferometers might have been better omitted, rather than discussed so briefly. Fourier transform spectroscopy, now a standard technique in the infrared, is mentioned by Smith and Thomson and by Welford, whereas it is dealt with in some depth by Möller. It is a little surprising that Smith and Thomson do not give the technique more space, as their book contains by far the greatest coverage of Fourier theory in general.

The basics of polarization are well treated in all three books, but Welford goes little beyond that. Smith and Thomson move on to give accounts of the transverse and longitudinal electro-optics effects, now used so widely in the optics laboratory, but only Möller proceeds to discuss one of the three basic calculation techniques for handling problems in polarized light: Jones calculus.

All of the authors have something to say on lasers, Smith and Thomson devoting a full chapter to the subject. Möller has surprisingly little on holography, considering the size of his book, whereas Welford gives more detail on holographic interferometry than one might have expected in a short, first-year undergraduate text. In the area of what might be described as modern optics, we have accounts of nonlinear optics in Möller, and Smith and Thomson (the former being the more detailed), integrated optics and acousto-optics in Möller, and optical lightguides in Welford. Möller has included a much more comprehensive treatment of detectors than is provided in the other books, and also has a chapter on scattering.

Which of the books is the most appropriate for a particular undergraduate reader depends on the content of the associated lecture course. The treatment of Fresnel diffraction and aspects of modern optics may be deciding factors. □

John Macdonald is a Lecturer in the Department of Physics, University of Reading, Whiteknights, Reading RG6 2AF, UK.

• *A Course in Mathematics for Students of Physics*, by Paul Bamberg and Shlomo Sternberg, is based on a course taught at Harvard intended for "students with an interest in physics who have a good grounding in one-variable calculus". Published by Cambridge University Press, the work is in two volumes; Vol. 1 is available. Vol. 2 is published in May.

## Grains to rock

Timothy Astin

**Applied Sedimentology.** By Richard C. Selley. Academic: 1988. Pp. 446. Hbk £39.50, \$79.50; pbk £16.95, \$34.

**Basics of Physical Stratigraphy and Sedimentology.** By William J. Fritz and Johnnie N. Moore. Wiley: 1988. Pp. 371. Hbk £40.50, \$45.65; pbk £12.95, \$11.85.

**Exercises in Physical Stratigraphy and Sedimentology.** By William J. Fritz and Johnnie N. Moore. Wiley: 1988. Pp. 221. Pbk £19.85, \$22.35.

**Techniques in Sedimentology.** Edited by Maurice Tucker. Blackwell Scientific: 1988. Pp. 394. Hbk £39.50, \$89.95; pbk £24.95, \$44.95.

THE study of sedimentary rocks can be loosely divided into two — sedimentology, which encompasses the origin, transport and deposition of rock-forming materials, the effects of depositional environments on rock-type distribution, and of alteration following burial (diagenesis); and stratigraphy, the definition and correlation of rock units, leading to sedimentary basin analysis. Parts of the field have grown rapidly during the past few years, notably diagenesis and more complex methods of basin analysis. Others — such as sediment transport and deposition, methods of analysis, and study of the general nature of sedimentary environments and elements of stratigraphy — are comparatively mature.

Three of the textbooks reviewed here confine themselves to well-trodden paths, reproducing much that is to be found elsewhere. The clearest example of reworking is *Applied Sedimentology*, which is a slightly modified, updated and improved version of Selley's *An Introduction to Sedimentology* (Academic, 1976), a second edition of which was published in 1982. There is a superficial freshness about the book, enhanced by the spuriously new title, but most of the content is the same. The retention of so much text written before 1975 has occasionally led to absurdities; for example, a newly inserted sentence, "Today the term geosyncline is obsolete", is followed by a lengthy discussion of geosyncline terminology from the old book.

Nonetheless, like its predecessor the 'new edition' has several virtues. All aspects of sedimentary rocks are covered, running from weathering and grain aggregate properties, through transport, depositional structures and environments and subsurface environments, to discussions of the main types of rocks and basins. The material in each section is simplified and easy to read, and the whole gives a helpful introduction for students —

as the title to the previous book suggested.

The companion volumes by Fritz and Moore take a very different approach to that adopted by Selley. The books are aimed specifically at students new to the subject, and are intended to provide the foundation for an introductory lecture and practical course. As the title indicates, the material covered in *Basics of Physical Stratigraphy and Sedimentology* is deliberately limited to stratigraphic principles, and sedimentary rock and structure description. The initial emphasis is on rock description, the account of which is followed by explanation of how transport and depositional processes lead to the structures seen in rocks. The natural next step, discussion of sedimentary environments, is left to other textbooks.

If *Basics* deals with the sort of material covered in lectures, *Exercises* has the material you would use in the following practical classes. Specific exercises are given, preceded by comprehensive notes on the topics covered, from rock description, through sedimentary structure analysis, to fence diagrams. I found *Exercises*, with its insight into how others teach students, the more appealing of the two books; and it made me wonder whether my own department ought to bind its lecture handouts and practicals together into a similar compilation.

In contrast to the other books, *Techniques in Sedimentology* breaks new ground. It is a response to the growth in use of methods of analysis of sedimentary rocks, especially as applied to diagenesis. The editor set out to cover all the common techniques, including collection of field data, analysis of grain size, optical microscopy, X-ray diffraction, scanning electron microscopy and chemical analysis. The battery of techniques concerned is still growing fast, and the instrumentation list could have been expanded. Obvious (and acknowledged) omissions are ways of measuring porosity and permeability, and all aspects of organic matter analysis (why most geologists consider organic matter as somehow separate from other parts of sedimentary rocks is a mystery). The account of field data collection deals with methods practised by the geologist with notebook, hammer, hand lens and compass, and does not sit easily with the rest of the contributions. Anything more adventurous (sediment peels, box coring and so on) has been deliberately left out.

The summaries of laboratory methods, especially as applied to diagenetic studies, are well done and will be invaluable aids to teaching. All of the techniques presented are likely to be used by the graduate student or applied sedimentologist, and they will find the book extremely useful. □

Timothy Astin is a Lecturer in the Postgraduate Research Institute for Sedimentology, University of Reading, Whiteknights, Reading RG6 2AB, UK.

# The local star

J.C. Brown

**Astrophysics of the Sun.** By Harold Zirin. Cambridge University Press: 1988. Pp. 433. Hbk £37.50, \$49.50; pbk £12.95, \$22.95.

**The Atmosphere of the Sun.** By C.J. Durrant. Adam Hilger: 1988. Pp. 168. £23.50. Distributed in the United States by Taylor & Francis, \$64.

BECAUSE of its proximity, the Sun inundates us with data and with theoretical problems. Few possess the diverse knowledge of physics required to understand and lucidly explain these problems; in consequence, most recent books on general solar physics have been collections of chapters by specialist authors, with all of the well-known drawbacks of such volumes. As the testbed of almost all of astrophysics, the physics of the Sun warrants more accessible, single-author treatment. Fortunately, there seems to be a resurgence in the publishing of such monographs; those by Zirin and by Durrant, although complementary in approach and content and contrasting in style, constitute admirable examples.

Zirin's *Astrophysics of the Sun* was spawned as a revision of his earlier *The Solar Atmosphere*, published by Gail & Blaisdell in 1966, but in its growth has gone far beyond that. Although the book still bears the hallmark of an observer and spectroscopist, the new volume embraces most of the prevailing ideas of topical interest in the physics of the entire Sun. Throughout, it also reveals the colourful personality of the author, in enthusing over results from his own team, and in the frequent philosophical and scientific asides which range from the profound to the flippant. The overall effect is to make for most enjoyable reading, although serious-minded readers may find the individualistic style irritating.

In particular, theorists may react against the familiar Zirin tirade against their cavalier use of data by pointing to his and his fellow observers' use of confusing terminology — such as 'faculae', 'plagettes' and 'network elements', all of which refer to a single phenomenon — and to mistakes in the theoretical parts of the text. Glaring examples are the conceptual, numerical and referential errors in the treatment of non-thermal *bremssstrahlung*.

Nonetheless — and despite Zirin's early disclaimer to theoretical expertise — all of the important general ideas in theoretical solar physics come across clearly and some, such as the prediction of the solar wind, are presented as succinctly and lucidly as I have seen anywhere. I was disappointed only in the chapter entitled "Questions", which is an anticlimactic

end to an otherwise fascinating book.

*The Atmosphere of the Sun* is much shorter and more restricted in scope than Zirin's book. Durrant concentrates largely on the formulation and explanation of the plasma and radiation theory needed to understand both the gross and the detailed structure of the outer layers of the Sun, complementing this with clear accounts of the relevant observations. By adopting such a didactic theory-based approach, however, he has succeeded in producing a highly coherent account of the physics of solar physics, although the final chapter on solar activity falls below the general standard. Few writers have

managed to present so well a mathematically rigorous development of complex physical problems, together with clear verbal explanations of the significance of the mathematics, as well as of the physics and the problems themselves.

Anyone venturing to read either of these books, whether as a research primer or for general scientific interest, will close it much better informed about the Sun. For those wishing to pursue a deeper interest in solar physics, I strongly recommend reading both. □

J.C. Brown is a Professor in the Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QW, UK.

# Unearthly ideas

B.E.J. Pagel

**Astrophysical Concepts, 2nd edn.** By Martin Harwit. Springer-Verlag: 1988. Pp. 626. DM98, £32, \$54.

**Physics of the Galaxy and Interstellar Matter.** By H. Scheffler and H. Elsässer. Springer-Verlag: 1988. Pp. 492. Hbk DM118, £38.50, \$69.50; pbk DM88, £29, \$49.50.

**Cauldrons in the Cosmos: Nuclear Astrophysics.** By Claus E. Rolfs and William S. Rodney. University of Chicago Press: 1988. Pp. 561. Hbk £59.95, \$74.95; pbk £27.95, \$34.95.

THESE days there is such a copious production of proceedings of conferences and advanced study institutes in astrophysics and cosmology that it is difficult to keep up. The supply of textbooks that either introduce the subject from first principles, or provide an in-depth survey of a particular field, is very much smaller. So the appearance (or in two cases reincarnation) of three major texts is a welcome event.

*Astrophysical Concepts* is the second edition of a book that I have been recommending (with reservations) to students for some time. First published in 1973 and based on courses for seniors and beginning graduate students at Cornell, it develops many physical ideas required in astrophysics from first principles. The author uses specific astronomical situations for illustration, rather than systematically running through astrophysics as such, although essential astronomical facts and terminology are summarized in an opening chapter and an appendix.

The great strength of the book lies in the lucidity and elegance with which chosen topics are quantitatively developed using elementary and clever arguments, instructive problems being distributed throughout, and in the sceptical spirit of enquiry that pervades the writing. The weakness lies in unevenness of treatment, accentua-

ted by absence of much serious revision in the main body of the text. A fresh chapter on cosmology has been added, and this gives a good description of recent research with some useful comments on the formation of galaxies, stars and the Solar System. But it is simply superposed on a lengthy treatment of issues current before 1973 that would have benefited from shortening in favour of newer results (for example the discovery of gravitational lenses, Hawking's work on the thermodynamics of black holes and re-evaluation of Dirac's 'large numbers' hypothesis in the light of the anthropic principle).

Too much attention is paid to the cranky concept of (potentially observable) tachyons. Here the author uses or belittles consequences of 'traditional' special relativity as it suits him, gets into a muddle as to whether the momentum of tachyons is real or imaginary, and claims that "the existence of tachyons would have important consequences in cosmology and in rapid communication across large distances" — much as one might claim that the existence of a perpetual motion machine would have important consequences in cosmology and in helping to solve the world's energy problems. This quirk can perhaps be excused on the grounds that it is bound to provoke the reader into thinking; there is rather less excuse for a few straightforward technical errors, such as treating the diffusion of Lyman- $\alpha$  photons in a nebula as an example of a random walk in space. A flawed masterpiece.

Galactic research is one of the oldest parts of astronomy, but one that never stays quiet for long. Successive advances in observational technique — optical, radio, cosmic ray, ultraviolet, near infrared, X-ray, gamma-ray, sub-millimetre and now (with the IRAS satellite) far infrared — have brought new insights and complications, and have stimulated theoretical developments in stellar dynamics and in the physics and chemistry of the interstellar medium. A remarkably complete and balanced survey of all these topics is provided by *Physics of the Galaxy*



and *Interstellar Matter*, a translation with minor updatings of *Bau und Physik der Galaxis* (1982).

A few of these subjects are covered as well or better in other books. But Schefler and Elsässer's unique contribution is to have presented the whole subject on a single canvas. They provide detailed, up-to-date information that is not accessible elsewhere in so digestible a form, covering such topics as positional astronomy, Galactic distribution of gas and dust, light scattering by dust, spiral wave theory and interstellar chemistry, as well as exposing the observational results and theoretical arguments with care and thoughtfulness. The translation is quite accurate and reads well apart from a few oddities of nomenclature (such as 'kernel' for 'nucleus', 'null-point' for 'origin' and 'nascent energy' for 'energy released'). The book is admirably illustrated with informative tables, diagrams and historical comments — it is good that this useful survey has now been made more accessible to English-speaking readers.

Nuclear astrophysics lies at the heart of our understanding of stellar evolution and the origin of the chemical elements. Neutrino observations have provided one of the subject's most dramatic triumphs in the case of Supernova 1987A, but also a long-standing dilemma in the low flux of high-energy neutrinos from the Sun; as Willy Fowler reminds us in his foreword to *Cauldrons in the Cosmos*, "We humans are mostly (90%) oxygen and carbon. We understand in a general way the chemistry and biology involved, but we certainly do not understand the nuclear astrophysics which produced the oxygen and carbon in our bodies".

What we do and do not understand is the subject of clear and authoritative descriptions by two active researchers in the field. The book covers the theory of nuclear reaction rates, the techniques of their measurement in the laboratory (which involves special difficulties due to the need to extrapolate to low energies) and the resulting characteristics of successive nuclear burning stages in stars. It provides also a lively and reasonably complete account of the astrophysical context in which they occur, and brief reviews of some current issues such as the solar neutrino problem, nuclear cosmochronology and isotopic anomalies in meteorites. Assuming a good background in nuclear physics and quantum mechanics (relevant parts of which can be obtained from D.D. Clayton's *Principles of Stellar Evolution and Nucleosynthesis*), one could not wish for a better account of the current state of knowledge (and uncertainty) about nuclear reactions in stars. □

B.E.J. Pagel is a Deputy Chief Scientific Officer at the Royal Greenwich Observatory, and Visiting Professor at the Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK.

## Initial prospects in spectroscopy

Peter K. Trumper

**Laboratory Guide to Proton NMR Spectroscopy.** By S.A. Richards. Blackwell Scientific: 1988. Pp.229. Pbk £10.50, \$26.50.

**Organic Spectroscopy.** By D.W. Brown, A.J. Floyd and M. Sainsbury. Wiley: 1988. Pp.250. Hbk £22.50, \$44.55; pbk £9.95, \$19.70.

**Interpreting Spectra of Organic Molecules.** By Thomas N. Sorrell. University Science Books, 20 Edgehill Road, Mill Valley, California 94941/Oxford University Press, UK: 1988. Pp.175. Pbk \$18, £15.

FEW things in the organic chemist's world have changed more dramatically in the past decade than has the arsenal of spectroscopic methods available. Nuclear magnetic resonance techniques, which less than eight years ago were known only to physical and analytical chemists, are now accessible to undergraduate students who pass a bar code pen over a few labels. In mass spectrometry, soft ionization allows large, polar and non-volatile molecules to display molecular ions. Fourier transform infrared spectroscopy, and hyphenated techniques such as GC-MS, GC-IR and even LC-MS, can now be carried out on affordable instruments. The time has come for undergraduate texts to reflect the widespread use of these new approaches.

Richards's *Laboratory Guide to Proton NMR Spectroscopy* is a work with a refreshingly casual yet informative style, offering a wealth of handy hints for obtaining and interpreting proton NMR spectra. There is a good discussion of sample preparation and solvent selection. The material presented is easy to understand, and meets the author's stated goal of guiding the relatively inexperienced chemist through the proton NMR spectrum in the real world.

Unfortunately, the real world here is a decade old. Suggestions about how to interpret NMR spectra are mainly directed to those obtained on continuous wave spectrometers; the comments on the use of the spectrometer are often irrelevant to the user of a modern instrument; only brief mention is made of FT techniques and of two-dimensional spectroscopy; and although NOE spectra are included, and subtraction techniques are discussed, the NOE spectrum shown uses integration instead. All in all, the author seems to have been reluctant to enter the modern era.

In their book, Brown, Floyd and Sainsbury direct themselves towards providing students with a training in the

basics of organic spectroscopy. The fundamental physical processes involved in the common techniques are nicely explained at a level appropriate for advanced undergraduates, while the exercises include relevant problems and provide IR, MS, proton and carbon NMR spectra, and ultraviolet-visible and combustion data. It would, though, have been good to see structures in the answer keys, as students are likely to have difficulty with some of the (proper) heterocyclic nomenclature.

The preface advertises recent advances in instrumental methods, but these are only mentioned in passing in the text. Soft-ionization techniques such as CI and FAB are described, but no spectra are given as examples (although a few appear in the exercises). Fourier transform NMR is briefly explained, but the examples and exercises use continuous-wave proton spectra almost exclusively. DEPT spectroscopy is the only multiple-pulse FT technique mentioned.

Sorrell's *Interpreting Spectra of Organic Molecules* is designed as a companion text for the usual introductory organic chemistry course. In that role, it succeeds very well. NMR, IR and mass spectra are discussed; the omission of ultraviolet-visible spectroscopy seems timely. The information is presented concisely, and the explanations of the physical processes involved in spectroscopy are pertinent and clear. The author effectively emphasizes problem-solving techniques. NMR spectra from a high-field FT spectrometer are taken as illustrations; this is a welcome change after the spectra from 60 MHz libraries used in the other two texts. The chapter on  $^{13}\text{C}$  NMR spectroscopy does well in explaining decoupling and the NOE, though no DEPT spectroscopy is covered.

Sorrell has done an exemplary job of showing how to integrate information from different spectroscopic techniques. His book will be entirely appropriate for introductory organic chemistry courses, and should fit in well with any laboratory programme. No problems or exercises are included in the text, but some are available as instructional software for an MS-DOS microcomputer. □

Peter K. Trumper is an Assistant Professor and Dana Faculty Fellow in the Department of Chemistry, Bowdoin College, Brunswick, Maine 04011, USA.

● Published by Pergamon towards the end of last year was *Problem Solving in Analytical Chemistry*, by Themistocles P. Hadjiioannou and colleagues at the universities of Athens and of Washington. The subtitle declares the volume to be a handbook containing over a thousand worked examples, problems and answers; intended uses for the volume are various, but the authors' general aim is to enable students to consolidate their theoretical knowledge by working through real problems.

# Excited states

Mike Ashfold

**Principles and Applications of Photochemistry.** By Richard P. Wayne. Oxford University Press: 1988. Pp.268. Hbk £30, \$55; pbk £12.50, \$24.95.

**Spectrophysics, 2nd edn.** By Anne P. Thorne. Chapman & Hall: 1988. Pp.390. Pbk £14.95, \$32.50.

Most of the energy received by the Earth comes from the Sun in the form of electromagnetic radiation. The visible wavelengths are vital to processes such as vision and photosynthesis, while the ultraviolet radiation is crucial to maintaining the mantle of stratospheric ozone — which, in turn, protects us from these potentially damaging shorter wavelengths. As Wayne rightly comments in his preface: "Photochemical processes are of the greatest importance to life on Earth". They are also of fundamental interest. Photochemistry is largely concerned with processes involving electronically excited atoms and molecules; these have different electronic structures from those of the ground state species with which we are most familiar and, as a consequence, may exhibit very different chemistry.

Wayne's book is a major update of his earlier text, *Photochemistry*, which was published by Butterworths in 1970. As before, it is very much aimed at undergraduates. Thus the book contains qualitative discussion of the absorption and emission of radiation, and of the various modes of decay exhibited by electronically excited molecules — for example, fluorescence, phosphorescence, non-radiative intramolecular processes including photodissociation, and intermolecular (collisional) energy transfer. Ample illustrative examples are cited in each case. Other chapters are devoted to the role of electronically excited species in synthetic organic chemistry, and to photochemical techniques. Here, as in several of the preceding chapters, one cannot help but appreciate the revolutionary effects that lasers are having on studies of photochemical change.

The book concludes with an overview of "photochemistry in action" — topics covered include atmospheric photochemistry, photosynthesis, vision, photography, polymer photochemistry, solar energy storage, optical brighteners and some potential medical applications. This last section overlaps considerably with *Light, Chemical Change and Life* (Open University Press, 1982), but Wayne's text is uniquely valuable because of its complete and balanced coverage both of the physical principles underlying photochemistry and of their applications.

The advent of lasers, especially the

tunable dye laser, has opened a new era of 'high resolution photochemistry' — an era in which gas-phase photochemical processes can be probed with quantum-state resolution. Such experiments have the potential to reveal the intimate dynamics of the photochemical event: they also serve to highlight the degree of convergence between the traditional fields of gas-phase photochemistry and spectroscopy. Anne Thorne's revamped *Spectrophysics* is distinctive in that (like the first edition) it provides a thorough grounding in the methods of experimental spectroscopy. It too is principally intended for undergraduates. It includes clear descrip-

tions of the *modus operandi*, and the relative strengths and limitations, of grating and prism spectrometers, of interferometers and Fourier transform methods, and of laser spectroscopic techniques, and concludes with a good introduction to plasma spectroscopy.

Both books are eminently readable and contain pleasingly few typographical errors. Both are enhanced by the extensive recommendations for further reading listed at the end of each chapter, and by their comprehensive indexes. □

Mike Ashfold is a Lecturer in the School of Chemistry, University of Bristol, Bristol BS8 1TS, UK.

# Little and large

John Mann

**Organic Chemistry, 4th edn.** By T. W. Graham Solomons. Wiley: 1988. Pp. 1,186 plus indexes. Hbk £48.75, \$54.95; pbk £18.95, \$17.95.

**Organic Chemistry, 2nd edn.** By G. Marc Loudon. Benjamin/Cummings: 1988. Pp. 1,259 plus indexes. \$58.25, £19.95.

**Intermediate Organic Chemistry.** By John C. Stowell. Wiley: 1988. Pp.268. £31.50, \$35.95.

**Experimental Organic Chemistry.** By Clark F. Most, Jr. Wiley: 1988. Pp. 586. £31.45, \$41.01.

"ORGANIC chemistry is boring and difficult. It's all facts and named reactions." This is a perennial complaint from undergraduates, and one that the authors of big, general textbooks try to address. Over the years such books have become increasingly colourful and fact-enriched; but they all still approach organic chemistry by means of functional groups.

The two on offer this year (Solomons and Loudon) are new editions from well-established authors, and both boast notable changes. Solomons has now adopted full-colour, but it is used more to highlight names and titles than for mechanisms (as in Peter Vollhardt's *Organic Chemistry*, published by W. H. Freeman and reviewed here last year). The molecular orbital pictures and the diagrams in the chapter on nucleic acids and protein biosynthesis are, however, very pretty.

As before, certain topics are designated 'special', and are dealt with in isolation. This is a drawback of this and many other textbooks, because it encourages an unhealthy compartmentalization of chemistry. Thus organosulphur chemistry merits five pages as a special topic in Solomons, whilst Loudon (using various shades of green and grey) introduces the subject throughout his book whenever comparisons with the corresponding carbon-oxygen compounds can best be

made. Similarly Loudon devotes a whole chapter to enolate chemistry and its applications in polyketide and fatty acid biosynthesis. Solomons relegates both topics to 'special' sections. The juxtaposition and comparison of similar chemistry is a positive feature of Loudon's book, and one reason why it makes more interesting reading.

Both volumes have been thoroughly updated, have more than adequate coverage of spectroscopic techniques, and contain extensive indexes and numerous problems (Solomons provides answers; Loudon does not, though he claims to include more problems than any other text). But do they present organic chemistry in a stimulating light? There has been a notable move away from lists of reactions towards understanding the chemistry through a mechanistic approach. Students will find both books informative but somewhat dull and ponderous.

One measure of readability is whether or not a book can be enjoyed in the bath. This would be physically impossible with Solomons and Loudon, both of which are weighty tomes, but John C. Stowell's *Intermediate Organic Chemistry* passes the test. The book is for students who have had an introductory course, and who then "pick up an issue of *Journal of Organic Chemistry* . . . and find the real world of the practising chemist to be . . . on a different level of understanding".

Much of the difficulty stems from chemical methodology or jargon that is beyond the standard textbooks, and Stowell's aim is to bridge this gap. He commences with an excellent guide to literature searching and to nomenclature, before plunging into functional group chemistry, carbon-carbon bond formation, and the planning of multi-step syntheses. The material is contemporary and fairly comprehensive, with research references up to 1985. But there is one big irritation. Stowell seldom provides any mention of the reasons for carrying out the chemistry. "Compound 6 is useful for the synthesis of hemlock alkaloids" is a typical statement; but how can the student tell,



when no structure of a hemlock alkaloid is provided, let alone even the briefest mention of their interesting toxicology.

Two chapters on reaction mechanisms follow, and the frontier molecular orbital approach receives a good coverage. There is an interesting chapter, full of practical hints about optimization of reactions through the use of temperature, pressure, polar solvents, crown ethers, solid supports and so on. The book concludes with a brief account of certain specialized techniques in NMR (COSY,  $^{13}\text{C}$ -spectra), and a tiny index. There are some appropriate problems, with literature references provided for those seeking the correct answers.

Overall, Stowell succeeds in his goal of helping people make the transition from basic organic chemistry into the realms of research. Once into the research literature, a student can always find something of interest.

Such forays also serve as a reminder that organic chemistry is above all an experimental science, yet there are few textbooks which deal with this side of the subject. Arthur Vogel's *Textbook of Practical Organic Chemistry*, published by Longman in 1978, is of course the classic,

and Clark F. Most's *Experimental Organic Chemistry* is rather like a slimmed-down version of that. (A new edition of Vogel is due later this year.) Most's book does have some appealing features of its own, such as the strong emphasis on safety, and some excellent tables on the chemical hazards and toxicities of common organic chemicals. It is, too, much easier to read than Vogel, and the introductory chapters on basic techniques, report writing, equipment, use of the research literature and so on could profitably be read by all undergraduates. The sections on spectroscopic methods are very good, but the chromatography is rather basic. When it comes to actual experiments, these are rather simple, and at least in Britain many of them would be done at school rather than at university.

Finally, there is a small section on qualitative analysis (Vogel has a very large section), and a modest index. The book is clearly written and illustrated, and with its strong emphasis on safety it should find a place alongside Vogel in all teaching laboratories. □

John Mann is a Reader in the Department of Chemistry, University of Reading, Whiteknights, Reading RG6 2AD, UK.

## Making a meal

J.A. McCleverty

**Advanced Inorganic Chemistry, 5th edn.** By F. Albert Cotton and Geoffrey Wilkinson. Wiley:1988. Pp.1,455. £21.95, \$44.95.

MOST inorganic chemists know the preceding four editions of this text well. So here I will not describe in detail the rich menu being offered, but will only compare it with its predecessor.

First, the repast is more substantial (59 extra pages). Second, each course is altered in size. Thus the multi-choiced *hors d'oeuvres* of the fourth edition have been replaced by a shorter, general survey (stereochemistry; simple bonding modes; general ligand behaviour;  $\pi$ -acceptor and hydrocarbon ligand types), reduced in length from 208 pages to 83; here I missed the comprehensive ligand review of the previous edition. The fish and meat courses follow tradition, although the coverage of main groups is expanded (by 34 per cent) and that of the transition metals is slightly abbreviated (by seven per cent). The special topic section is, as before, the varied and delectable crowning confection of the book, and has been reorganized and lengthened by 62 pages.

This new edition is a treat, and no self-respecting inorganic chemist should deny him or herself. It is up to date (with

references to 1987), and full of relevant examples to modern petrochemical-based processes. However, it is not appropriate as a basic student text, except for those who intend to take inorganic chemistry very seriously (others should turn to the same authors' *Basic Inorganic Chemistry*). For instructors, it is a mine of information, but may only be fully appreciated with previous editions and other related volumes — the Mrs Beeton and Robert Carrier of the inorganic world — to hand. Cotton and Wilkinson avoid much theoretical discussion, pointing out the transitory nature of such punditry, and use the space released from the previous editions to add new facts. In *The Chemistry of the Elements* Greenwood and Earnshaw take a broadly similar view, although more from a main-group than a topic-orientated perspective, while in *Inorganic Chemistry* Purcell and Kotz adopt the typically North American theoretical approach but sacrifice facts.

The fifth edition is adequately, if somewhat idiosyncratically referenced. The appendices are of limited use, and I am not convinced they are necessary. But although the book is not comprehensive (there is hardly a mention of solid-state chemistry or of the influence of the subject on materials science), it is an invaluable guide. Together with the volumes mentioned above, it will grace my kitchen for many years to come. □

J.A. McCleverty is a Professor in the Department of Chemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.

## The matter in suspension

John Gregory

**Basic Principles of Colloid Science.** By D.H. Everett. Royal Society of Chemistry: 1988. Pp.243. Pbk £9.95, \$19.50.

COLLOID science is not well endowed with introductory texts (or any texts for that matter), so the appearance of Professor Everett's book is very welcome. The author has set out to survey the field from a fairly fundamental standpoint, but without assuming too much in the way of background knowledge. His aim was for the book to be accessible to readers at about A-level chemistry standard, although there is much here that will interest more advanced students, and those in industry who are working with colloidal systems but don't have any formal training in the subject. There are many examples given of technological aspects.

On the whole, the principles are adequately covered, although, in places, I have doubts about the approach taken. There is frequent recourse to formal thermodynamic argument, which is not always appropriate at this level of exposition and which sometimes gets in the way of a clear conceptual understanding of the phenomena described. However, several difficult concepts, such as dilatancy, are put over very clearly, with good explanatory text and helpful illustrations.

Given the central role played by surface charge (at least in aqueous colloids), surprisingly little space is devoted to electrical double layers and electrical interaction between particles. The Stern layer and specific counterion adsorption are hardly mentioned, yet these concepts help greatly to understand colloid stability.

There is a potential source of confusion over terminology. In his introductory chapter Professor Everett distinguishes between 'coagulation' and 'flocculation', and yet in Chapter 9 the terms are used virtually interchangeably. Again, although the index contains separate entries for 'heterocoagulation' and 'heteroflocculation', the references are to essentially the same phenomenon.

These and a few other slight blemishes should not detract from the value of the book. It provides a thorough overview of modern colloid science, with a nice balance between the fundamental and the applied aspects. It will serve as a good introduction at several levels, and will be greatly appreciated by teachers and practitioners of the subject. □

John Gregory is Reader in Water Chemistry, Department of Civil and Municipal Engineering, University College London, Gower Street, London WC1E 6BT, UK.

# All to do with genes

Graham Boulnois

**Genetic Elements in *Escherichia coli*.** By Peter Smith-Kleary. Macmillan, London: 1988. Pp.198. Hbk £25; pbk £9.95. To be published in the United States by Guilford, New York; price has yet to be decided.

**Genetic Engineering.** By J.G. Williams and R.K. Patient. IRL: 1988. Pp.72. Pbk £5.95, \$11.95.

**Gene Structure and Transcription.** By T. Beebee and J. Burke. IRL:1988. Pp.77. Pbk £5.95, \$11.95.

**Genetic Engineering: An Introduction to Gene Analysis and Exploitation in Eukaryotes.** By S.M. Kingsman and A.J. Kingsman. Blackwell Scientific: 1988. Pp.522. Hbk £39.50, \$89; pbk £19.50, \$43.95.

THE techniques of molecular biology were largely developed for the genetic analysis of *E. coli* and its phages and plasmids, but rapidly found application in studies of the structure and function of eukaryotic genomes. The result has been not only an explosion in our understanding of gene organization and expression in higher organisms, but also the appearance of ever more powerful methods with application in prokaryotes and eukaryotes. Students therefore ought to acquire a solid background in both prokaryotic and eukaryotic molecular biology, yet the two branches are often treated as independent entities. In their different ways (but with the possible exception of *Gene Structure and Transcription*) these books continue to promote the divide. That said, both teacher and student will find much to interest them here.

The biological basis of recombinant DNA technology is well documented by Smith-Kleary. He discusses in detail the often-forgotten classical *E. coli* genetics which played such a crucial role in defining systems — such as the *lac* operon, phage lambda and plasmids — that are now the cornerstone of modern molecular biology. The combination of classical bacterial genetics and modern molecular genetics to be found in the book will fascinate those brought up on bacterial genetics. The problem-solving approach to the topic will be welcomed by many teachers, but may be received with less enthusiasm among 'technologically minded' students. The power and elegance of the classical genetics complements and contrasts with the more straightforward molecular genetics, and this is a good text for those wanting to look back upon the development of genetic engineering. But I wonder whether it will have a

place in a modern molecular biology course, where teaching time is at a premium.

The IRL "In Focus" series continues with *Genetic Engineering* and *Gene Structure and Transcription*. Both maintain the excellent record of these volumes, which are cheap, up-to-date, and provide clear and concise views of rapidly expanding fields. Of the two I found Beebee and Burke's *Gene Structure and Transcription* more interesting and informative — it provides a wealth of information on gene expression in both prokaryotes and eukaryotes, and well illustrates the power of the analytical methods concerned. Second- and third-year undergraduates would do well to acquire it.

In *Genetic Engineering*, Williams and Patient take a rather traditional approach (if this is possible in such a new field) to the description of how to clone and characterize genes. Much of the material can be found in other, more comprehensive texts, but the new approaches to cloning and analysis are made especially comprehensible with the aid of clear illustrations. To complement more biologically based texts, and as a revision aid, this book will be useful.

As an illustration of how genetic engineering can be exploited to probe basic

problems in biology, and of how it finds application in biotechnology, the text by Kingsman and Kingsman is excellent. The authors go beyond the isolation and characterization of genes and take the view, quite rightly, that manipulation of cloned genes and their reintroduction into cells is at the heart of things. The techniques, their application to biological systems as diverse as yeast and mammals, and their use in the analysis of processes ranging from photosynthesis to carcinogenesis, are clearly described and critically assessed. The figures, too, are good. This text will be widely used by final-year students, and many graduate students will find it a useful source of background information.

Given that there are now a great many textbooks on molecular biology and genetic engineering in competition for the student's money, in a sense it is sad to report that none of these books can be ignored. Each of the first three has its specialist niche, while Kingsman and Kingsman will find a place in supporting more general courses. □

Graham Boulnois is a Lister Institute—Jenner Research Fellow and Reader in the Department of Microbiology, University of Leicester, Medical Sciences Building, Leicester LE1 9HN, UK.

## Figuring it out

Simon Hunt

**Complement.** By S.K.A. Law and K.B.M. Reid. IRL:1988. Pp.72. Pbk £5.95, \$11.95.

**Lymphokines.** By A.S. Hamblin. IRL:1988. Pp.71. Pbk £5.95, \$11.95.

**Immune Recognition.** By M.J. Owen and J.R. Lamb. IRL:1988. Pp.73. Pbk £5.95, \$11.95.

**Essential Immunogenetics.** By A.R. Williamson and M.W. Turner. Blackwell Scientific:1988. Pp.432. Pbk £19.50, \$46.

DAVID Male, editor of the immunology titles in IRL's "In Focus" series, is a formidable optician. Between them, he and his authors write clearly, present the truth without gloss and bamboozle rarely. The first three of the series succeed authoritatively in making some fast-moving areas accessible and bang up to date. The format is attractive, with fresh two-colour diagrams, and occasional spectacular electron micrographs or full-colour graphics, and there are plenty of references. The amount of experimental description is well balanced. All of the authors finish by peering enthusiastically into the future.

*Complement* could easily have been the customary catalogue, but Law and Reid's account is a revelation, integrating the

kick-off of the two pathways, the control mechanisms and the fascinating molecular homologies (C1q with collagen; the serine proteases; the perforins with terminal attack-proteins; the short consensus repeats) with the consequences for pathological processes. You might think that every possible representation of the pathways has been tried, but I have not seen it done better than in Fig. 1.1 of this book.

Likewise, in *Lymphokines* Hamblin uses her first diagram for crisp scene-setting. For each of the interleukins 1 to 6, gamma interferon and the cytokines acting on the various haemopoietic compartments, she covers methodology, biochemistry, some genetics and their *in vitro* properties, ending with some possible clinical applications. *Immune Recognition*, however, was perhaps allowed to stray too far. In this case Fig. 1.1 shows MHC restriction in T-cell-mediated cytotoxicity; Owen and Lamb then shoot through immunoglobulin, MHC antigen and T-cell-receptor structure, function and gene diversification, and just squeeze in a section on thymocyte maturation. It's all a bit fast and dense.

Picking off areas one by one is of course what students and their tutors usually do, but with the eventual penalties of some overlap and of loss of broad perspective. So there is still a place for a good general textbook. Might *Essential Immunogenetics* do the job? The authors seek to make their subject more approachable to new-



comers who may find it formidable — this is supposedly an entry-level book. However, it belies its title.

There are several defects. The authors fail to anticipate the reader's likely difficulties in grasping the cock-eyed nomenclatures with which the subject has grown up. The book is also full of mistakes. Again, Fig. 1.1 sets the tone. In it, Thy-1 is incorrectly shown as having a transmembrane hydrophobic peptide; a depiction of the correct lipid anchor doesn't arrive until page 298, though even here Fig. 12.10 dithers. Out of 16 figures in Chapter 7 that have been taken from the original literature, seven have mistakes, four of which are so serious as to make the diagrams incomprehensible. Errors in the text include malapropisms (for example, 'syngeneic' for 'syntenic', page 266) and contradictions with the figures. No reader of a textbook should have to double-guess its authors to such an extent.

Finally, the book is out of date. Its structure rests mainly on a historical

narrative, which frequently doesn't connect with the present. The introduction hints that the pregnancy through which its authors laboured was like that of a marsupial suffering delayed implantation, and it shows. The references in Chapters 5 and 6 ("Classical Genetics of Immunoglobulins" and "Antibody Diversity") stop at 1981; those in Chapters 9 and 10, on the genetics of the MHC, include only one later than 1983. So there are gaping structural omissions.

Is there really even a separate discipline of immunogenetics that requires an undergraduate textbook? This one, with its excursions towards the scattered destinations of immune proteins, genes, cells and diseases, raises unfortunate doubts whether there is. The publishers ought in this case to reassess the balance between their own reputation and what they owe to their authors. □

Simon Hunt is a Lecturer in the Dunn School of Pathology, University of Oxford, OX1 3RE, UK, and a Fellow of Keble College.

much effort as the writing of the text. Several of them show reconstructions of extinct forms disporting themselves in naturalistic, if idyllic, settings; and even the skeletons are caught in the act of behaving.

The occasional error has crept in (*Allocebus* is not the same size as *Microcebus*; lemurids and cheirogaleids — as the illustrations show — do not have three incisors), but these are commendably few for such a wide-ranging work. One caveat, though: it is nice to see body weights quoted for extinct primates, but because most are perforce extrapolated from dental dimensions the confidence with which they are presented may be a trifle excessive.

Whereas Fleagle's book fills a long-standing need for a comprehensive and up-to-date introductory text in its field, *Human Biology* is the third edition of an old favourite first published in 1964. Commendably, the authors have resisted the dismal trend towards simplification that has afflicted introductions to physical anthropology over the past couple of decades; for this reason — as well as, perhaps, the price — the book may, alas, prove in practice more welcome to those preparing comprehensive courses in this eclectic field than to those taking them.

As ever, authoritativeness is assured by the division of responsibility for the four sections (roughly, the human evolution, variation, growth, and adaptability of the subtitle) among four specialists. This new edition is dedicated to the memory of the two of the original four authors (J. S. Weiner and Nigel Barnicot) who have, sadly, died since the last edition appeared. Their original co-authors, as well as their successors, David Pilbeam and Paul Baker, have produced new or revised texts that are informative and thoughtful, and are fully worthy of that memory. □

Ian Tattersall is Curator in the Department of Anthropology, American Museum of Natural History, New York, New York 10024, USA.

## Ordered order

Ian Tattersall

**Primate Adaptation and Evolution.** By John G. Fleagle. Academic: 1988. Pp. 486. \$39.95, £28.

**Human Biology: An Introduction to Human Evolution, Variation, Growth, and Adaptability, 3rd edn.** By G. A. Harrison, J. M. Tanner, D. R. Pilbeam and P. T. Baker. Oxford University Press: 1988. Pp. 568. Hbk £40, \$70; pbk £20, \$35.

EVEN a superficial comparison of John Fleagle's new introduction to the primates with its predecessor, Le Gros Clark's *The Antecedents of Man*, dramatically illustrates how much primatology has changed over the three decades separating the two books. The comparative anatomical foundation on which *Antecedents* was based has expanded enormously and has been given an important functional component; a vast quantity of behavioural and ecological information has been amassed; and our understanding of primate systematics has been revolutionized, both practically and theoretically.

Each of these developments is reflected in Fleagle's book, which is a *tour de force* as well as a *tour d'horizon* of the primate order. The preliminaries are dealt with briskly in three short chapters, "Adaptation, Evolution, and Systematics", "The Primate Body" and "Primate Life". Most of the rest of the book is composed of accounts of the main groups of primates, the living followed by the extinct.

For the living forms, Fleagle briefly surveys behaviour, ecology, morphology and

relationships to the level of the genus; the fossil chapters, inevitably, lean more to morphology, though ecology is not ignored. Intervening between the living and fossil sections is a chapter entitled "Primate Adaptation", an interesting miscellany running from body-size effects to the locomotor correlates of ecology. A classification to genus level is given in the first chapter, but neophytes might have benefited from a fourth introductory chapter with an overview of primate systematics and diversity.

One of the most attractive aspects of the book is the way in which extinct primates are presented not simply as teeth and bones but as living, behaving creatures whose lives as a whole we can hope to interpret. This emphasis is complemented well by the illustrations, the preparation of which must have cost the author — let alone his talented artists — at least as



Thereby hangs a tail — a troop of vervet monkeys in woodland savannah.

From Primate Adaptation and Evolution

# Two complete brains

Hendrik Van der Loos

**Neurobiology, 2nd edn.** By Gordon M. Shepherd. Oxford University Press: 1988. Pp. 689. Hbk £35, \$47.95; pbk £19.50, \$32.95.

**The Human Brain.** By Paul Glees. Cambridge University Press: 1988. Pp. 204. £32.50, \$59.50.

IN discussing two books in the same review, a compare-and-contrast approach seems inevitable. Here, that is perhaps not a bad thing because both are textbooks which deal with the whole brain, including that of human beings. Both are also new editions; Shepherd's book was first published in 1983, while Glees's appeared in German in 1968. (Or was it 1971? The answer is not made clear.)

Single-author volumes that cover as much ground as these two are an increasing rarity, and we could do with more of them. But while *Neurobiology* demands admiration for its completeness, *The Human Brain* leaves one uneasy. Far too many important areas have been left out — why, for example, is there nothing on Geschwind and brain asymmetry; on Mountcastle and the posterior parietal cortex; on Hökfelt and multiple transmitters; on Aguayo and regeneration? The omissions from Shepherd's book are on a much smaller scale. In the fine chapter on developmental neurobiology, the remodelling of neurons, including their axons, might perhaps have received more attention, as might the factors that shape the central homeomorphic representations of peripheral sensory sheets. The same holds, in the chapter entitled "Sensory Systems", for the projections from cortex to thalamus and to the nuclei of termination.

As to the books' heuristic value, the figures in Shepherd complement the text well and both are clear. One exception is the very first diagram, with its legend and surrounding text, which fails to explain fully the link between "nervous organization and behavior", held by the author to be a key issue. Again, *The Human Brain* suffers in comparison: several figures do not make the point they are supposed to make; others show items not mentioned in the text; yet others do not seem to make

any point at all. Some statements in the book are impenetrable. For example (here I paraphrase) the five brain vesicles are said to "contain the imprints of the genetic programmes which emerge in birds after hatching"; gonadal activity ceases early (in aging) and thus, Glees suggests, neuro-secretory nerve cells stop functioning before those neurons that generate impulses only; the high degree of orderliness seen in the cerebral cortex is not found at lower brain levels "where neurons are lumped together in scattered groups referred to as nuclei" — here I suggest that the student should get hold of a microscope and have a quick look at retina and cerebellar cortex for themselves.

The two authors take a comparative approach, but a totally different one. For most tasks carried out by the brain, Shepherd shows how invertebrates and vertebrates go about solving them — an interesting way to teach the subject (for example in "Communication and Speech", one is told about insect song and birdsong). Glees instead begins with the brain of the lamprey and, by a set of edicts, converts it into the human brain. In places the body leads the way ("The forearm . . . becomes reconstructed . . ."); elsewhere, the brain appears to be the prime mover of evolutionary change; and elsewhere still, the

senses are in charge. All of these factors may have a role in evolution. But to my mind the time is not ripe to integrate them, and Glees certainly does not try.

In standard of production, both volumes have defects. The paper on which Shepherd's is printed could have been better. In my copy, text from the verso side of a page can sometimes be read in mirror image on the recto. And some pages consist of two pieces glued together (fortunately with generous overlap), which diverts one into idle reflections on how books are made. Glees's seems to have escaped normal quality control; figures commonly lack scale bars and the reproduction of many of them is substandard. The text is so riddled with misprints that not even German words and names have been spared.

Each book contains a wealth of data and insight, and *aficionados* of neuroscience will appreciate that. Novices will come away from *Neurobiology* enriched — though perhaps overwhelmed (there are over 600 pages of text). But although such readers may emerge from *The Human Brain* with a heightened sense of curiosity, they will also emerge confused. □

Hendrik Van der Loos is Director of the Institute of Anatomy, University of Lausanne, rue du Bugnon 9, 1005 Lausanne, Switzerland.

# Worlds together

Alan Brafield

**The Invertebrates: A New Synthesis.**

By R.S.K. Barnes, P. Calow and P.J.W. Olive. Blackwell Scientific: 1988. Pp. 582. Hbk £32.50, \$57.53; pbk £15.95, \$28.24.

TWENTY years ago, in the preface to his *Invertebrate Structure and Function*, E.J.W. Barrington wrote that "the business of animals is to stay alive until they have reproduced themselves, and the business of zoologists is to try to understand how they do it". Such an understanding requires appreciation of two very different things: the enormous diversity of animals, and the unifying features of comparative physiology and behaviour. Barrington's book was a milestone in the comparative functional approach as against the phylum-by-phylum treatment. He was ahead of his time. In the genre, only G.S. Carter's *A General Zoology of the Invertebrates* had appeared earlier, and few such books have been written since. With the functional approach, rather than detailed anatomy and systematics, now being in favour, the time is ripe for a synthesis of the two philosophies.

Here it is. *The Invertebrates: A New Synthesis* canters through the groups and then considers comparative physiology

and behaviour. It is an attempt at the best of both worlds, and it works. Some provocative opinions expressed in the first part (for example that molluscs are acoelomate and simply jumped-up flatworms which have gained a shell and a radula), though possibly nearer the truth than the conventional view, should have been explained at greater length, because students will not find such heresies in other books. My only other serious criticism is that the inclusion of full classificatory trees listing all the orders (48 in the case of the molluscs) is inappropriate here — particularly as there are numerous drawings of common genera, to show the range of form, in which the animals are unfortunately not identified.

The second part of the book is thought-provoking, up-to-date and very well illustrated. The chapter on respiration is too short; this is a wide-ranging subject which deserves more than half the space afforded to feeding. But the one on biomechanics and locomotion is interesting and comprehensive; that on defence stimulating; that on reproduction and life cycles scholarly; that on embryology and development informative; and that on control systems (by D.W. Golding) superb. Students will like this book. It deserves to succeed. □

Alan Brafield is a Senior Lecturer in the Division of Biosphere Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

• A new contender in the reference-book stakes is *Dictionary of Biology*, by W.G. Hale and J.P. Margham, both of Liverpool Polytechnic. Running from A to zymogen granule, and covering aspects of medicine and agriculture, as well as biology, the entries have been written primarily with school students and undergraduates in mind. Publisher is Collins.



## Reviews on IMMUNOASSAY TECHNOLOGY

*Edited by S. B. Pal*

### Volume 1.

October 1988 £37.50 256pp

234x156mm 0 333 452852

Apolipoprotein quantitation by ELISA: Technical aspects and clinical applications - J. B. Bury & M. Y. Rosseneu

Recent progress in the quantitation of phagocytosis - T. W. Jungi

Monoclonal antibodies directed against human immunoglobulins: preparation and evaluation procedures - J. J. Haaijman et al

Conjugation of haptens and macromolecules to phycobiliprotein for application in fluorescence immunoassay - N. Monji & A. Castro

Nicotine enzyme immunoassay - A. Castro & N. Monji

Applications of human monoclonal antibodies in non-isotopic immunoassays - S. A. Gaffan and M. C. Glassy

Adaptation of enzyme labelled immunoassay for measurement of biosynthesis of human - L. Morrison et al

Immunofluorescent methods for the assay of cytoskeleton antibodies in human sera - D. Zauli et al

The use of Western blot procedures in the analysis of *Herpes simplex* virus proteins - J. Kuhn et al

Non-isotopic studies of the TSH receptor - N. R. Farid and G. Fahraeus-Van Ree

Index

### Volume 2.

October 1988 £37.50 216pp

234x156mm 0 333 46849X

Human monoclonal antibodies produced by Epstein-Barr virus immortalized cell lines: technical and theoretical principles - M. Steinitz

Sensitive enzyme and erythro immunoassays for detection of human pregnancy - S. K. Gupta

Automated methods to test polymorphonuclear leucocyte function: - P. De Paoli & G. F. Santini

The Epstein-Barr virus hybridoma technique for production of human monoclonal antibodies - R. Burioni et al

Applications of liposomes in non-isotopic immunoassays: a review - N. Monji et al

Establishment of a particle concentration fluorescence immunoassay for the measurement of human IgG in culture supernatants - M. B. Compas & Y. G. Alevy

Establishment of a very sensitive avidin-biotin ELISA for the measurement of human IgG - Y. G. Alevy & C. M. Blynn

Methodology, limitations and uses of antigen-binding assays in the analysis of T-cell receptor-antigen interactions - B. E. Elliott et al

Spin immunoassays - C. C. Curtain & L. M. Gordon

Enzyme-linked immunoassays for detection of anti-drug antibodies - J. W. Coleman

IgE measurement in serum and culture supernatant - C. Gaveriaux & F. Looor

Index

*Order these books through your bookseller or send your order to:*

**Richard Gedye, Macmillan Press Ltd, Houndmills, Basingstoke RG21 2XS, UK.**

## Index to books reviewed

- Advanced Inorganic Chemistry, 5th edn  
by F. A. Cotton & G. Wilkinson 182
- Applied Sedimentology  
by R. C. Selley 178
- Astrophysical Concepts, 2nd edn  
by M. Harwit 179
- Astrophysics of the Sun  
by H. Zirin 179
- The Atmosphere of the Sun  
by C. J. Durrant 179
- Basic Principles of Colloid Science  
by D. H. Everett 182
- Basic Solid State Chemistry  
by A. R. West 175
- Basics of Physical Stratigraphy and Sedimentology  
by W. J. Fritz & J. N. Moore 178
- Cauldrons in the Cosmos: Nuclear Astrophysics  
by C. E. Rolfs & W. S. Rodney 179
- Complement  
by S. K. A. Law & K. B. M. Reid 183
- Electrons in Solids: An Introductory Survey, 2nd edn  
by R. H. Bube 175
- Essential Immunogenetics  
by A. R. Williamson & M. W. Turner 183
- Exercises in Physical Stratigraphy and Sedimentology  
by W. J. Fritz & J. N. Moore 178
- Experimental Organic Chemistry  
by C. F. Most, Jr 181
- Experimental Physics: Modern Methods  
by R. A. Dunlap 177
- Fundamentals of Solid State Physics  
by J. R. Christman 175
- Gene Structure and Transcription  
by T. Beebe & J. Burke 183
- General Relativity: A Guide to Its Consequences for Gravity and Cosmology  
by J. L. Martin 176
- Genetic Elements in *Escherichia coli*  
by P. Smith-Kleary 183
- Genetic Engineering: An Introduction to Gene Analysis and Exploitation in Eukaryotes  
by S. M. Kingsman & A. J. Kingsman 183
- Genetic Engineering  
by J. G. Williams & R. K. Patient 183
- Human Biology  
by G. A. Harrison et al. 184
- The Human Brain by P. Gleees 185
- Immune Recognition  
by M. J. Owen & J. R. Lamb 183
- Intermediate Organic Chemistry  
by J. C. Stowell 181
- Interpreting Spectra of Organic Molecules  
by T. N. Sorrell 180

- An Introduction to Solid State Diffusion  
by R. J. Borg & G. J. Dienes 175
- The Invertebrates: A New Synthesis  
by R. S. K. Barnes et al. 185
- Laboratory Guide to Proton NMR Spectroscopy  
by S. A. Richards 180
- Lectures on the Electrical Properties of Materials, 4th edn  
by L. Solymar & D. Walsh 175
- Lymphokines  
by A. S. Hamblin 183
- Neurobiology, 2nd edn  
by G. M. Shepherd 185
- Optics by K. D. Möller 177
- Optics, 2nd edn by F. G. Smith & J. H. Thomson 177
- Optics, 3rd edn  
by W. T. Welford 177
- Organic Chemistry, 2nd edn  
by G. M. Loudon 181
- Organic Chemistry, 4th edn  
by T. W. G. Solomons 181
- Organic Spectroscopy  
by D. W. Brown et al. 180
- Physics of the Galaxy and Interstellar Matter  
by H. Scheffler & H. Elsässer 179
- Primate Adaptation and Evolution  
by J. G. Fleagle 184
- Principles and Applications of Photochemistry  
by R. P. Wayne 181
- Spectrophysics, 2nd edn  
by A. P. Thorne 181
- Statistical Mechanics  
by B. K. Agarwal & M. Eisner 175
- Statistical Physics  
by A. Guénault 175
- Statistical Physics, 2nd edn  
by F. Mandl 175
- Techniques in Sedimentology  
M. Tucker, ed. 178

**nature**  
is available in  
microform  
from University  
Microfilms  
International.

☐ Please send information about these titles:

Name \_\_\_\_\_

Company/Institution \_\_\_\_\_

Address \_\_\_\_\_

State \_\_\_\_\_ Zip \_\_\_\_\_

Phone ( ) \_\_\_\_\_

Call toll-free 800-521-3044. Or mail inquiry to:  
University Microfilms International, 300 North  
Zeeb Road, Ann Arbor, MI 48106.